BOSTON STUDIES IN THE PHILOSOPHY OF SCIENCE

VOLUME XLVIII

MODELS

SYNTHESE LIBRARY

STUDIES IN EPISTEMOLOGY,

LOGIC, METHODOLOGY, AND PHILOSOPHY OF SCIENCE

Managing Editor:

JAAKKO HINTIKKA, *Academy of Finland, Stanford University, and Florida State University*

Editors:

VOLUME 129

MARX W. WARTOFSKY

MODELS

Representation and the Scientific Understanding

For Carol

EDITORIAL PREFACE

Marx Wartofsky has been working for many years within an unusual confluence of philosophical problems. He brings to these intersecting problems his comprehensive intelligence, at once imaginative and rigorous, analytic and historical. He is a philosopher's philosopher, but also Everyman's.

Wartofsky is philosopher of the natural and the social sciences, of perception, esthetics and the creative arts, of the 18th century French and the 19th century Germans, of politics and morality, of the methods and morals of medicine, and it is plain, of all human existence. To a colleague, he seems Jack-of-all-philosophical-trades, and master of them too. The reader soon will learn that Wartofsky is a genial, lucid and relaxed philosophical companion, deeply serious but without noticeable anxiety.

I need not highlight these selected epistemological papers gathered as, and about, *Models*, since Wartofsky's own introductory remarks are helpful and stimulating in that respect. I need only, after 21 years of friendship and collaboration with him, warn the reader to beware of how profound and provocative these papers will show themselves to be beneath their good-humored and swiftly-flowing surface.

And I must publicly note the pleasure with which I welcome Marx Wartofsky's volume to our *Boston Studies*.

Boston University   R.S.C.
Center for the Philosophy and History of Science
September 1979

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

xi

INTRODUCTION

The essays in this volume were not written for this volume. Each had its own rationale, in its own time and place. Many were written for the occasion: a symposium here, a *Festschrift* there, a collection or an invitation to speak, a sudden insight just as suddenly caught up in the pains and frustrations of making articulate sense of it. In short, there is no claim to a deep and conscious project, of which these essays were the stages along the way. I cannot even put the burden upon the cunning of reason, working silently and tacitly through me. I would like to. But what I have to offer is something less and something more.

Several years ago (in January 1974), on the occasion of a series of four invited lectures at M.I.T., I attempted the systematic construction of a thesis concerning the genesis and historical evolution of what I called "modes of cognitive praxis", in particular, science and art. This was an attempt to show that human knowledge itself has a history. By this I meant not simply that what we know has a history, or that there is a historical development of ideas or theories; but that the nature of knowing, of cognitive acquisition itself, changes historically; *how* we know changes with changes in our modes of social and technological practice, with changes in our forms of social organization. In effect, I argued that what we take knowledge to be is itself the subject of an historical evolution. The claim was based on an analysis of what I took to be the crucial feature of human cognitive practice, namely the ability to make representations. This I traced to the primary production of artifacts – in the first place, tools and weapons, but more broadly, in good Aristotelian fashion, anything which human beings create by the transformation of nature and of themselves: thus, also language, forms of social organization and interaction, techniques of production, skills. The production of such artifacts for use, I argued, was at the same time the production of representations, in that such artifacts not only have a use, but also are understood as representing the mode of activity in which they are used, or the mode of their own production. Thus, spears and axes are not only made for the sake of hunting and cutting, but at the same time represent both the method of their manufacture and the

activities of hunting animals or chopping wood. The argument I offered there was that such external representations, grounded in the primary activities of the production and reproduction of species life, are the precondition of so-called internal representations, that is, for the reflective activities of imagination, thought, conscious purpose, which are the evolved characteristics of human cognition.

The aim was nothing so grand as an account of how we get from the amoeba to Einstein; on the other hand, it wasn't exactly a modest project either. I intended to show how the highly developed forms of representation, in science and art – i.e. scientific theories and representational works in painting and literature – could be seen to have their genesis in those modes of representation which emerge simultaneously with our primary productive, social and linguistic practice. I called the project 'historical epistemology'. The lectures, enlarged since then beyond the original four, are not yet published, though I have since tried ⋅to elaborate the epistemological thesis, in various versions, and with ramified applications; and I have tried also to give it further historical and anthropological grounding.

This long aside is intended to suggest something *less* than the cunning of reason at work in the present essays: namely, the discovery that, in retrospect, many of my earlier essays came into focus as undeliberate and yet preliminary forays in search of a thesis. History is retrospective reconstruction, and lends a seductive cogency and coherence to the past, as all of us who have heard the flapping wings of the owl of Minerva well know. But in fact, it was the initial articulation of the historical epistemology thesis that first led me to see in a surprising proportion of my earlier occasional pieces a kind of bumbling prolegomenon to a more systematic construction.

Why is this something *more* than the cunning of reason, however? Simply because it yields no agency to any suprahistorical Idea, concedes to no higher rationality at work than my own, and is ready to make do with error as no accident, but as part of what it takes, and costs, to keep going. This is certainly not to claim that there aren't strong influences upon my own thought, hidden borrowings, continuities of which I will never be aware, historical and contemporary antecedents, some of which I know very well, others of which only others will know and recognize.

What is crucial in the essays in this volume is the notion of representation, and in particular, the role and the nature of the model, in the natural sciences, in theories of perception and cognition, and in art.

Philosophy itself I see as the systematic construction and logical/critical analysis of such models, in their fullest elaboration as formal structures, as ontological claims about the nature of things – worlds, societies, persons, actions, thought itself – , and as heuristic constructions which suggest how we should proceed in structuring our understanding of the world and of ourselves. What I can attempt, in this introductory essay, is no more than a programmatic sketch of the notion of representation, and of the use of models, as the framework within which, retrospectively, the essays in this volume come to be seen as fragmentary adumbrations of a more general thesis.

In science, as in much of art, human knowledge is achieved by means of representation. By contrast with nonhuman animals, human beings create the means of their own cognition. That is to say, we create cognitive artifacts which not only go beyond the biologically evolved and genetically inherited modes of perceptual and cognitive activity, but which radically alter the very nature of learning and which demarcate human knowledge from animal intelligence. The cognitive artifacts we create are models: representations to ourselves of what we do, of what we want, and of what we hope for. The model is not, therefore, simply a reflection or a copy of some state of affairs, but beyond this, a putative mode of action, a representation of prospective practice, or of acquired modes of action.

The more ordinary sense of what a model is takes it as a selective or abstractive duplication of some aspects of the world. In effect, the model is taken to be a construction in which we organize symbols of our experience or of our thought in such a way that we effect a systematic representation of this experience or thought, as a means of understanding it, or of explaining it to others. But such a view presupposes what it means for something to be a representation, and avoids precisely the crucial question of how it is that we can come to represent. The issue, therefore, is "how is representation possible?" And of course, this question requires an analysis of what a representation is; or more precisely, what it is we do, when we represent.

The cognitive artifacts which we create, and by means of which we acquire knowledge, are various. If we are to start from a consideration of the most highly developed of such artifacts, then we would probably choose scientific theories as the paradigmatic cases of such representation. We should also include systems of logic or of mathematics, and the forms of representation in literature and art. These are the most fully

evolved modes of representation which we have achieved. The question concerning each of these remains "how is a scientific representation, or a mathematical representation possible? How is a literary or a pictorial representation possible?" In short, two systematic issues present themselves, one analytic, the other genetic. The analytic issue concerns how we are to understand the nature and function of a representation, in any of these modes. The genetic issue concerns how we have been able to generate the very activity of representing itself, how we have come to create cognitive artifacts in the first place. For if it is true, as I claim, that human beings distinctively create the means of their cognitive acquisition, the question is, by what means, and in the course of what activity does this take place.

The genetic argument is barely touched upon in the essays in this volume, though the suggestions as to how one could approach it are made. Representation, and the construction of models, is seen to be based in the very character of human practice, or action. The two fundamental forms of our activity are the making of things, and social interaction. These activities, especially in their integral relation to each other, are the requisites for our existence, day-to-day, and from one generation to the next. The making of things is essentially related to the production and use of tools, as social interaction is essentially related to the production and use of language. Tools and language, then, become the basic artifacts by means of which the human species differentiates itself from its animal forebears; and it is therefore in an analysis of these basic artifacts that a theory of the genesis of representation needs to be developed.

Aristotle had already introduced the distinction between the making of things, or production, and social interaction. The first activity, *poiesis*, had as its object things made; the second *praxis*, had as its object other persons or social interaction. What remained undeveloped in Aristotle, because of the separation in his theory between practical production and social or civic life, was the essentially social nature of production itself. The human making of things, in its fundamental historical forms, involves social interaction. It is a social practice. The capacity for social communication is therefore a requisite for production itself. So too, therefore, are the forms of social organization within which this production takes place. Language and social organization are thus not only conditions for the life-sustaining activities of the human species, but give human productive activity its special character.

Marx had the insight into this relation between production and social interaction, between making and doing. In his conception of this integral process of human activity, he characterized it as *praxis*, (combining the senses of both terms, *poiesis* and *praxis*, in this broader conception, but retaining the emphasis on social interaction which the original term had in Aristotle's usage.) Dewey, too, saw the systematic connection between communication and production, in characterizing language as the "tool of tools". By this, he meant that everything that concerns the making and use of tools involves communication, language, precisely in that human production is a social activity.

Thus, after quoting Franz Boas as saying "The two outer traits in which the distinction between the minds of animals and man finds expression are the existence of organized articulate speech in man and the use of utensils of varied application"[1], Dewey goes on to say:

. . . at every point appliances and application, utensils and uses, are bound up with directions, suggestions and records made possible by speech; what has been said about the role of tools is subject to a condition supplied by language, the tool of tools.[2]

My view of representation, however, is not that all representation is linguistic, or that 'internal representation' is somehow 'inner speech'. It is rather different and more complex than such a view, and the essays in this volume do not go very far in either posing this problem or resolving it. Minimally, however, I may suggest that representation is already involved in the very character of non-linguistic artifacts, insofar as they are themselves symbols, or are parts of systems of symbols, in which the artifacts themselves present meanings, intentions, relations, and come to represent to us the modes of practice involved in their production and use. This external representational form, the embodied model, so to speak, is not yet a linguistic artifact, though as a kind of physical inscription, it presages the kinds of signs which become elements of a constructed language, and insofar, provides the genesis for pictorial, hieroglyphic and alphabetic forms of language, i.e. for diagrammatic, or written linguistic forms. The detachment of the representational sign from the primary artifact, the invention of conventional signs, of formal systems, of mathematical, pictorial and linguistic symbol systems goes far beyond the original interpretation of the tool as itself the symbol of the mode of action by which it is made, or in which it is used.

The connection between language, (in particular, vocal speech), and the making of things, is therefore a subject still to be broached and

investigated. Marx's striking aphorism, "Language is practical consciousness", requires the elaboration that it is also social consciousness, that is, the medium of communication and expression in the contexts of social interaction as well as production; and clearly, not all of human activity is production. But all production is essentially social; and in this sense, there is no more 'private production' than there is 'private language'.

   These speculative and programmatic suggestions aside, it remains clear that language is crucial in theoretical activity, and that both natural and constructed languages are central to theory-formation in the sciences; but so too are the non-linguistic modes of representation, what one may characterize as the pictorial or diagrammatic modes, or in their mathematical form, the geometric modes. Models thus have their genesis in a number of such modalities, and differ in their character and use thereby. What models have in common, on this view, is that they are deliberately constructed representational artifacts; and as such, they are, in all their modes, the means by which human consciousness presents itself with its own objects, i.e. in which it becomes self-consciousness. By this, I do not mean simply consciousness of a 'self', but rather consciousness of one's own (or of the species' own) activity, of the practice which connects us with and helps to constitute the world, and through which we come to understand it, to know it; in effect, to propose, to conjecture, to entertain and to test what may be *true* about it. Models are proffered truths. To proffer truth is the human means of acquiring knowledge. In this sense, cognitive acquisition, human learning is essentially mediated by representation. It is what makes theory possible.

   What is at issue, in representation, then, is the distinctively human way in which we acquire knowledge. Strictly speaking, one may say that there is no human knowledge without representation; or more radically still, that there is no knowledge without representation. This is to claim that knowledge, suitably defined, is a distinctively human achievement, and that it is to be qualitatively distinguished from animal intelligence and learning, from the ways in which non-human animals acquire learned modes of behavior, or inherit structural modalities of perception, feeling and motor activity. Before developing such a view, it is clear that certain conceptual and terminological distinctions need to be made.

   First, if we are to limit the concept of cognition, or of knowledge to human learning, this imposes a rather strict requirement with respect to

any talk about animal 'cognition', namely, that such talk is metaphorical, and essentially anthropomorphic. It reads back into non-human learning a category which is, strictly speaking, alien to it. And yet, there is no question that non-human animals learn, that they are conscious or aware of what they perceive, that they solve problems, and that they acquire skills. More crucially, non-human animals can be puzzled, can make mistakes, can act in order to satisfy needs, can be frustrated or satisfied. Whatever reduction of these states the anti-cognitivist behaviorists may want to effect, by way of interpreting such 'cognitive' states and capacities as no more than patterns of behavioral response shaped by schedules of reinforcement, it seems to me that animal activity betrays search, choice, goal-orientation, affect, in a way which clearly marks it as conscious activity, and not mere response-mechanism. And yet, I want to reserve the terms 'cognition' and 'knowledge' for use in the exclusive contexts which involve representation; and to claim that animals neither make nor use representations in their conscious activity, (though they may use what *we* make as, or take to be representations).

Second, if the term *representation* is to be used in a way which demarcates human from non-human animal consciousness and activity, then its use in any talk about 'internal representation' in non-human animals, or indeed, any talk of representation in machines, e.g. in computers, or by means of mechanical or electronic or chemical reproduction, is likewise metaphorical, anthropomorphic and parasitical for its meaning upon the human activity of representing.

Such conceptual and terminological *caveats* are not intended to cramp the language, nor to condemn the extended usage of 'knowing' and 'representing' which pervades our talk about animal learning and machine computation and reproduction. Rather, the intention is to recognize that such usages are metaphorical, rather than mistaken; and mistaken only when it is forgotten that they are metaphorical and anthropomorphic. In fact, an enlightened anthropomorphism is perhaps the most powerful heuristic framework for inquiry into animal behavior and computational process. For example, it is useful and fruitful to think of animals as developing 'cognitive maps', in Tolman's sense; and to think of machines as 'adding', 'solving equations' or 'comparing and assessing probabilities'. I would not even want to propose that we redescribe this by saying "machines do not add or solve equations but do something *like* what we do when *we* add or solve equations". It is not

necessary that computers be said to do 'something like what we do', to avoid the anthropomorphism, because they may be 'doing' *nothing at all* 'like what we do' in these contexts. It is enough that we can use the *model* of 'adding' or 'solving' *to represent to ourselves*, and to understand better what it is the machines are 'doing', or how they operate (since a machine may be 'doing' nothing at all, but simply moving from one state to another, in accordance with a program). Conversely, it may be that we come to understand aspects of what we do, in computing or solving problems, by representing it in a model of machine operations, stripped of all (or nearly all) anthropomorphic metaphor. But again, the model need not be taken in any sense as an account of what we *do*, when we compute or reason. Similarly, the imputation of cognitive processes to animals may be heuristically powerful in suggesting problems for research, or in providing a model of animal behavior which aids us in explaining it, or understanding it. For these are precisely the functions which models have, as I will argue. What is required is that, in these cases at least, we keep in mind the metaphorical nature of such models, and remember to distinguish the model from what it is a model of.

But this raises, at the outset, a serious problem. Are models, therefore, no more than calculated fictions, known to be false, but used because instrumentally, and epistemologically fruitful? Are models ever 'true', i.e. is a model ever a true representation of its object? This gets us into what it is for something to be a representation in the first place. And here the crucial issue is joined.

Either some things are representations of others by virtue of some intrinsic or relational properties the representation has with respect to what it represents; or its *being* a representation depends on something else. The 'something else' is what I will argue for. The argument, in a nutshell, is that:

(1) *Anything* (in the strongest and most unqualified sense of 'anything') can be a representation of anything else. Therefore, there are no intrinsic or relational properties which mark one thing off as a representation of something else; or (as Goodman has pointed out) everything has infinitely many properties in common with everything else, and so anything can be taken as a representation of anything else in terms of some of these shared properties.

(2) It is *we* who constitute something as a representation of something else. It is essential to something's being a representation, therefore, that it be *taken* to be one.

(3) From (1) and (2) it follows that a representation is whatever is taken to be a representation; that representing is something *we do*, and that nothing *is* a representation except insofar as we construct or construe it as one; and in this, it is precisely the representation we make it, or take it to be.

This argument shifts the emphasis from talking about what representations *are* to the consideration of the human activity of *representing*, since in effect representations are nothing but what they are made to be, in and by this activity. To put this somewhat differently, representations are, paradigmatically, intentional objects. They come to be what they are, are sustained or maintained as such, and are exhaustively describable in terms of our own intentions. Further, since the function of a representation is to stand for something beyond itself, not simply to denote it, but to present the thing represented to us in some way in which we come to understand it, representation essentially involves both reference and meaning. But referring here is itself again something *we* do, and is an intentional activity. Nothing is, intrinsically, referential, nor do any two things stand in the relation of referring apart from our constitution of the relation as intending the reference. This is not to say that there are *two different* relations, representing and referring; but rather that in representing, we also refer; or that reference is a constitutive aspect of representation, and that intending to represent, or in our taking something to be a representation of something else, we also intend to refer. It is, in effect, part of what we mean to be doing, when we are representing.

Talk about intentional objects typically conjures up ontologies of mental entities which exist 'in the mind' or 'for the mind', and are somehow to be distinguished from proper spatio-temporal physical objects. Representing, insofar as it is a contemplative or reflective activity of the imagination, i.e. in those contexts in which we entertain thoughts or call up or create mental pictures of a perceptual sort, in terms of visual or auditory imagery, for example – is, in my view, the latest and perhaps most complex mode of representing activity. It is in this sense that most talk about intentional objects takes place, whether these are taken to be mathematical objects, or concepts, or perceptual images, i.e. objects of memory or of imagination which are formed in terms of the modalities of sense-perception. But I would argue that such 'mental' objects, or 'internal representations' are derivative, and have their genesis in our primary activity of representing, in which we take external

things, – most typically, what we also designate as physical objects – as representations. Moreover, I take our *making* of representations to be, in the first place, the actual praxis of creating concrete objects-in-the-world, *as* representations; or of taking the made objects as representational.

These preliminary considerations on the nature of representation still say nothing about its historicity. They do, however, suggest how an historical epistemology may be developed, and how it is to be more than a history of ideas, (or even of ideas about ideas). If, in fact, our modes of cognitive practice change with changes in our modes of production, of social organization, of technology and technique, then the connection between cognition and action, between theoretical and applied practice, between consciousness and conduct, has to be shown. Otherwise, we simply have an account of correlation, without any mediation between the correlated contexts. The one-sided manifestos of classic philosophical materialism and idealism – 'being determines consciousness' or 'consciousness determines being' – leave out of account the crucial question: How? Or at best, the 'determination' is construed either as the essentially passive 'reflection' in consciousness of the 'world', or even of our 'praxis', in an outmoded and epistemologically anomalous 'copy theory'; or else the 'world' becomes 'my idea', or even '*the* Idea' in its (dialectical) othersidedness: nothing but the outward product of a self-differentiating activity of consciousness itself. Between an inert epiphenomenalism, and a hyperactive creationism, we are left with two unacceptable theories of mind; and in consequence, two distorted theories of science.

It seems to me that a theory of representation, based in the practical activities of the making of things, and of social interaction and communication, bears the promise of a non-reductive historical account of the growth of knowledge, and of the evolution of the cognitive means by which such knowledge grows. It is within such a framework that I think the essays in this volume may be read in a more coherent way than they were originally intended.

Thus, on this retrospective reconstruction, the essays here fall into certain groups, and may be seen not as themselves *part* of a project, but rather as aspects of the formulation of a project, as part of its own history, so to speak. The initial group concerns, most directly, the notion of a model, and specifically, the notion of a theoretical model in the sciences. The second group deals with the active role which

representation and the use of models plays in shaping our perception and cognition, and in serving as a heuristic guide to our theoretical or practical activity. Here, it is suggested that picturing, in art, and constructing theoretical models, in science, are both modes of a more basic activity of representation; and that our own perceptual and cognitive understanding of the world is in large part shaped and changed by the representational artifacts we ourselves create. We are, in effect, the products of our own activity, in this way; we transform our own perceptual and cognitive modes, our ways of seeing and of understanding, by means of the representations we make.

In a sense, this is a sort of historicized Kantianism. Theoretical artifacts, in the sciences, and pictorial or literary artifacts, in the arts, constitute the *a priori* forms of our perception and cognition. But contrary to the ahistorical and essentialist traditional forms of Kantianism, I propose instead that it is we who create and transform these *a priori* structures. Thus, they are neither the unchanging transcendental structures of the understanding, nor only the biologically evolved *a priori* structures which emerge in species evolution (as, for example, Piaget and the evolutionary epistemologists suggest). Piaget's dynamic, or genetic structuralism is important here, of course. His dictum, ''no genesis without structure, no structure without genesis'', suggests the dialectical interplay of the practical emergence and transformation *of* structures with the shaping of our experience and thought *by* structures. But the domain of this genesis I take to be the context of our social, cultural and scientific practice, and not that of biological species-evolution alone. Historical epistemology begins where evolutionary epistemology lets off. Human perception and cognition have a history, beyond species evolution, and it is a history we ourselves have made, in the course of our socio-cultural practice. In a sense, then, our ways of knowing are themselves artifacts which we ourselves have created and changed, using the raw materials of our biological inheritance. And the suggestion, in a number of the essays here, is that the crucial means by which this self-transcending construction of our perceptual and cognitive modes takes place is representation.

A third group of essays is principally historical, and deals with the philosophical conceptions which shape our understanding of the world, and of our own perceptual and cognitive capacities. It is a matter of circumstance that the figures dealt with are Spinoza, Hume and Diderot. The rationale for the inclusion of such studies in the history of

philosophy is given in the essay 'Metaphysics as Heuristic for Science', where I argued (in 1965, when it should already long have been a commonplace, but wasn't yet) that scientific theory construction has in fact, always been profoundly influenced by metaphysics, and that much of modern metaphysics was indeed in the service of theory construction in the sciences. (I would argue that this is so, as well, for Plato and Aristotle, and for much of Greek thought generally, but that is a longer story, part of which I began to formulate in the first part of my book, *Conceptual Foundations of Scientific Thought*, (1968), and which is omitted here.)

In 'Metaphysics as Heuristic for Science', I give a critique of those views which, initially, do show the relevance of metaphysics for science, namely, the views of Popper, Agassi and Kuhn. In a later essay ('The Relation of the History of Science to the Philosophy of Science', 1974), I try to trace the history of the critique (and self-critique) of positivist and logical empiricist philosophy of science, and the growth of the historical understanding of science, and of its metaphysical component, which I had sketched in the 1965 essay.

A final group of essays touch on the notion of the model in art, and also on the notion of art as itself a model of practice. The relevance of these considerations to those concerning the role of models in science is not explicit in these essays. But they do bear on the broader conception of representation which is the focus of this collection.

One last thing needs to be said concerning the way in which we come to know by way of our modes of representation. The talk about models, and about representation in general usually imputes an epistemology in which the knowing subject confronts a surrogate object of knowledge – the model – as a representation of the 'external world'. The model models the world; or if not the world, then society; or abstract relations among mathematical or logical entities; or else, it represents the human as an object for human cognition. However, in the conception of the model offered here, it is always at the same time a representation of the knowing subject as well: every model proposes a certain relation to the world, or to its object, and implicates the maker or user of the model in this relationship. We can therefore always read back or reconstruct the modeller from the model itself. It is someone who stands to the world, or to others, in the relation which the model proposes. In this sense, all modes of representation can become themselves modes of self-knowledge as well. The self-conceptions of a community of investigators is as much

a feature of a given theoretical model in the sciences, as is the conception of the world which the model presents. The model therefore also contains the peculiar physiognomy, the character and belief system of the cognitive subject, precisely in that it presents a world which is the world (or a putative world) *for* that subject. The knowing subject thus comes to know him/herself in knowing the world; and thus too, comes to change him/herself in changing that world. Not a new idea at all; but worth remembering in the context of a theory of representation.

   One may put this yet another way, by means of an example: An intriguing question arises from the ubiquitousness of handprints in paleolithic art: How shall we interpret this perhaps most ancient of all the visual artifacts of the human race? The simplest answer seems to be that such prints are deliberate marks of presence: 'I was here'. That is the simplest reading we ourselves can give of such a sign, for the handprint entails a hand pressed against a wall, with the express intention of leaving its imprint there. But read further: the handprint records a gesture, an action, an intention. Friday's footprint in the sand was not made for the sake of recording his presence for Robinson Crusoe; yet it did, unintentionally. It did so because it unambiguously designated its maker, and the act of his walking in that place. Crusoe brought with him enough of human understanding to read back from the print the presence of another man like himself. The hunter learns to read the animal's track, the spoor, and in this he carries into human cunning and the human imagination a skill already emergent in the hunting animals. Sheer perceptual recognition of prey and predator is rooted more deeply still in the neurophysiology of vision, of smell, mapped genetically into the distance-receptors, into the very cells of animals. But a handprint is a human being's deliberate representation of him/herself, recognizable as one's own, and more: as like that of others like him or herself. The marks by means of which a species represents itself *to* itself are thus distinctive. They involve more than the innate, or even acquired species-recognition which is involved in the perceptual systems of all animals. Such marks involve the conscious purpose of self-representation, the conscious imperative to know oneself, or to make oneself known to others which entails an act of recognition that is distinctively human. Here I will only suggest that this element in our representational activity functions not only in graffiti or in self-portraiture, but in our theoretical models in science as well. We become theoretical in knowing ourselves to be theoretical; and we know ourselves to be theoretical in the very act of

creating theoretical artifacts. To allege that theory is the graffiti of the intellect is not to diminish theory, in my view, bur rather to acknowledge the importance, in our quest for objective knowledge, of the human imperative to make one's presence known, and to be recognized by the fruit of one's labor.

   This is a dangerous way to introduce a volume of one's essays. But these are my models, the ways I have come to understand the question of representation itself, and to arrive at my present formulation of the problem. They permit the critical reflection, revision and rejection which are required if one is to proceed further.

NOTES

[1] Franz Boas, *The Mind of Primitive Man*, New York: Macmillan, 1938, p. 98.
[1] John Dewey, *Experience and Nature*, La Salle, Ill., 1929, 2nd ed., p. 140.

THE MODEL MUDDLE: PROPOSALS FOR AN IMMODEST REALISM

[1966]

The aim of this essay is to give a frankly representationalist account of the nature and role of models in science. To this end, I propose to collapse the distinctions between models, theories, analogies, and to take all of these, and more besides, as species of the genus representation; and to take representation in the most direct sense of image or copy. The realism which the thesis involves is that which takes the objects, events or processes which are represented in a model as material objects, or events and processes of a material world. The disaster of devolving upon a crude and naive realism will be avoided, hopefully, by avoiding a crude and naive concept of representation – that is, a simple copy theory. The tactic then is to enrich the concept of representation in such a way that it can accomodate a fairly sophisticated range of scientific models. Among these, there are those which seem furthest removed from any notion of representation; and if it can be shown that these too may be interpreted as representations, presumably this will operate recursively over the more explicitly and frankly representational types.

   The problem to which this program addresses itself I have called the model muddle. The symptom of the muddle is the proliferation of strange and unrelated entities which come to be called models. Thus, 'model' is used for the straightforward mechanical model – the Tinkertoy construction – which is used in classroom demonstrations and in museums; as well, for the theoretical-construct in physics or in psychology which has its embodiment only in mathematical or verbal inscriptions or utterances, but which purports to give a factually true description of some part of the real world outside of language; and equally, for the mapping of some uninterpreted formal system on some interpretation or embodiment of it, which may or may not have any purported reference outside language, or outside the system of inscriptional shapes of a logic. On the other hand, in an attempt to order the proliferation of loose and ad hoc usages, models have been distinguished from theories, and from analogies (or from 'mere analogies') on the grounds of the different degrees of commitment one

has towards them as (in Kemeny's phrase) "intended factually true descriptions," or the different degrees of systematicity of structure or the differences in scope which each exhibits. Thus Achinstein has recently argued[1] for such distinctions, stressing that analogies, on the one hand, are to be characterized mainly by "physical similarities between analogue and theoretical object", and not simply by formal similarities of structure; that models are something else, again: apparently "a set of assumptions or postulates describing certain physical objects or phenomena . . ."; these in turn are distinguished from the pictures or diagrams which may be helpful in explaining or illustrating features of the model, but are not themselves models; and models and theories are distinct in that

To propose something as a *model* of (an) x is to suggest it as a way of representing x which provides at least some approximation of the actual situation; moreover, it is to admit the possibility of alternative representations useful for different purposes. To propose something as a *theory* of (an) x, on the other hand, is to suggest that x's are governed by such and such principles, not just that it is useful for certain purposes to represent x's as governed by these principles or that such principles approximate those which actually obtain.[2]

The difference, then, would seem to be in the way in which one *takes* a set of assumptions, the seriousness with which one entertains them, or the intention of the reference as only approximative or useful, or as a committed cognitive claim to truth.

At the outset, I would like to abandon such distinctions, expressed as "that's not a model, that's a theory", or "this is only an ad hoc analogy, shouldn't be mistaken for a full-fledged theory", not because such distinctions may not justifiably be made, on some specification of the differing meanings of these terms, or the different uses one may empirically discover as a descriptive linguist. Rather, I want to suggest a universal relation which all such entities exhibit – one in which they are the same rather than different. I have argued (in another paper[3]) for a typology of models, ranging them hierarchically with respect to the degree of existential commitment which each type suggests. On this ordering, ad hoc analogies do exhibit the weakest commitment, and in increasing strength, I have ranked computational models (mathematical models whose purported function is only as 'inference-machines', or as frameworks for ordering data), mechanical models, hypothetical models of the 'as if' sort, models which are regarded as approximative only, those which function as conjectures, and finally, those which command

existential commitment at the limit of our rational belief: the ones we feel
sure about, or those the belief in which cannot plausibly be controverted.
I do not therefore eschew distinctions. But it seems to me that
distinctions among degrees of belief or existential commitments need to
be supported by an analysis of the ways in which models of all these types
represent: therefore, not simply an analysis of their structural properties,
but of the relation between such properties and the purported reference
of the models.

The model muddle is one which I take to be a semantic
muddle – ultimately, an ontological muddle – regarding on the one hand
the status of the entities we call models, and on the other, the status of
what it is they are models of. In much of model-talk, models inhabit a
limbo between worlds. On the one hand, they are not citizens of the
blood-and-guts world of real objects and processes; or at best have only a
derived citizenship by way of their reference to such a world. On the
other hand, they are denied full equality in the cognitive world of
purported truths, assigned only the function of instruments of such
cognition: crutches, aids to the imagination, inference-machines,
heuristic devices, data-ordering frameworks and whatnot. When Russell
opted for logical constructions over inferred entities, he sought to avoid
the problem. But it was never clear what the status of logical
constructions was. If one reverted to a Platonic realism, then the logical
construction didn't avoid the problem, but only raised it in a different
way. The ten years' discussion instigated by McQuorquodale and Meehl
among psychological theorists, and deriving from issues raised by
Tolman and Hull, concerning hypothetical entities and intervening
variables seems to have ended in a muddle worse than the rather simple
one with which it began. The working scientist, in view of this, is prone
to take an attitude of 'don't-bother-me-now-I'm-working' towards all
this 'clarification', and to go on proliferating models of every sort and
description, with jolly abandon. In general, philosophers of science,
during that period of shamefaced ontology which characterized logical
empiricism's sway, came to afford to models the status of the Metics in
ancient Athens: necessary for the commerce and well-being of the state,
but aliens nevertheless, tolerated only for their usefulness. I am
suggesting that this instrumentalist immigration policy needs to be
radically overhauled.

The way to begin is to be clear about a trivial truth: models exist, or in
the deviant usage of philosophers: 'There are such things as models'.

There is an additional trivial truth, which may strike some people as shocking: anything can be a model of anything else! This is to say no more than that between any two things in the universe there is some property they both share, there is some relation which they bear to each other. But from this *embarras de richesses*, there are clearly only some things which we choose to sort out as models of some other things; and this, by specification of the relevant respects in which such things are related. More than this, merely to pick two things out as being related in some relevant respect is not yet to use one as a model of the other. Our own cognitive activity enters here, to take one as representing the other, and here there are constraints on what may or may not be made into a model. Here, we may resort to the simplest case of what the notorious plain man takes a model to be: I think he takes it to be an abstractive representation of some object or state of affairs. That is, he takes the model to exhibit only certain properties – the ones relevant to his needs and interests in using the model – of the object modelled. Thus, a model of the solar system, whether hanging from the ceiling in a Planetarium, or diagrammed two-dimensionally on a piece of paper, or described in the utterances of some speaker, represents a state of affairs in some abstractive version: *Some properties* of the model resemble, or are like, or image, or mirror, or stand in for some properties of the solar system. It is a redundant caveat to urge: 'The solar system isn't *really* like that', because, in the first place no one makes this mistake in its gross form – (no one seeing a model of the solar system ever thinks of planets as *being* ping-pong balls, or of orbits as being ellipses made of wire) – and in the relevant respect, the model *is* a claim that the solar system is precisely like that. The business about degrees of approximation is also a shabby complaint, if one keeps in mind a simple constraint on models, which we may take as a definition (or part of one), or as a convention: nothing which is a model is to be taken as a model of itself, nor of something identical with it. The reason for the constraint derives from the formal analysis of models. Inverse to the ordinary view of models as abstractive representations of some object or state of affairs, logicians speak of models as the interpretations or embodiments of some formal calculus, in which the relation of isomorphism (more strictly, homomorphism) holds between the structure of the formal system and that of its interpretation. In this sense, if one were to conceive of an axiomatized formal system which could be mapped on, for example, the mathematical structure of classical mechanics, classical mechanics would

be such an interpretation or embodiment. In such logical analysis (e.g. in Tarski, Kemeny, Suppes, Carnap, among others) the uninterpreted formal system is designated a *theory*, and its interpretation, a model of that theory. At the limit, where there is a unique model of the theory, or where for a given domain there is a unique theory, the relation is one of isomorphism. Thus Apostel can say, in this sense, that "The aim of science is to construct a calculus for which reality is the only model".[4] This raises a question. Since, on formal analysis, the relation of isomorphism is reflexive, and anything is isomorphic with itself, presumably anything is a model of itself if this structural or formal condition is taken strictly by itself. The condition of 'interpretation' or 'embodiment', involving as it does an intensional theory of reference seems to obviate this reflexive case, though only on some independent grounds on which something cannot be taken as an interpretation or an embodiment of itself. Most of the definitions and notions of model include some such constraint, but not all. In a weak sense, one may enforce the constraint by stating that at the limit, the case of anything being a model of itself is trivial. But Rosenblueth and Wiener[5] are willing to go all the way ("The best material model for a cat is another, or preferably the same cat"[6]). Skinner gives a similar example in *Verbal Behavior*:

... to the extent that responses resemble stimuli, responses related to similar stimuli will themselves be similar. Models have this property. We report a state of affairs most completely by reconstructing it – by building an exact duplicate. Such behavior is verbal according to our definition, since a model is built and used because of its effects upon 'listeners'. . . . Pointing to an object is a variation on model-building. A man may say, *I never go out without carrying my* . . . and finish by displaying an automatic drawn from his belt.[7]

On Skinner's account, then, the act of displaying an object or state of affairs makes it in effect a model of itself, in this context of verbal behavior.

If we enforce the constraint, that nothing be taken as a model of itself, or of something identical with it, then the relation of modelling is asymmetrical in some way. It would be nice if we were able to assign this asymmetry to the nature of the entities which stand to each other in the relation of *model of*. The universe could then be neatly arranged into at least a partial order of all those things which could serve as models and those which could not; or perhaps of all those things which could serve as models of anything else but themselves and the things which serve as

models of them. But the essential feature of the modelling relationship which makes such an account unfeasible is that, although it is the case that anything may be taken as a model of anything else, it is *being taken as a model* which makes an actual out of a potential model; and every case of *being taken as a model* involves a restriction with respect to *relevant properties*. Not everything has the relevant properties which permit it to be taken as a model of something else, if relevance has to do with *our purposes* in taking something as a model. If, indeed, relevance is restricted to the domain of human intentions and human discourse, then it is we who sort out, from the plethora of available properties, those which will establish, between two entities, the relation of modelling. Thus, although the modelling relation appears to be dyadic, it is in effect triadic: anything can be taken as a model of anything else if and only if we can sort out the relevant respects in which one entity is like another, the relevant properties which both have in common. Lewis Carroll poses the riddle that Alice cannot answer: "Why is a raven like a writing desk?"; but if Alice had been able to pick out relevant respects, she could have gone on answering until the end of the book. The limits on modeling are therefore the limits of our conceptions of what respects are relevant to what purposes, and not intrinsic limits which we can foreknow or prescribe for any two entities. We may then formulate the modeling relation triadically in this way:

(1) $M(S, x, y)$

where S takes x as a model of y. But this doesn't ensure asymmetry since S may very well turn around and also take y as a model of x. The constraints of taking the model (or in the inverse logical usage, the theory) as linguistic and the reference of the model (or the interpretation or embodiment of the theory) as extralinguistic, as is sometimes suggested, seems unnecessarily restrictive. The asymmetry of the relation comes rather from a practical interpretation of it; namely, that the model as a representation is not known to be as rich in the properties which are relevant as is the object or state of affairs to be modelled. In one sense, if we take our constraint seriously, then it cannot be as rich in properties – i.e. equally rich in the same properties – for then it would be identical with its object. It may be *richer* in properties, but these would then not be ones relevant to its object; it wouldn't possess them, and so the model couldn't be taken to represent them in any way. It follows then that the model has to be less rich in the range of relevant properties than its

object or reference. In the case of isomorphism, it would seem that the one unique representation would be as rich in its representation as anything we would know to take as a relevant property of the object or state of affairs. The 'One True Theory' would encompass all that could be encompassed in the way of representation. But at this limit, we are led to the Platonic identity of Being and Knowing, or the 'Absolute Identity' of Schelling's idealism which Hegel characterized as "the night in which all cows are black". Barring such gnostic excess, the *adequatio rei in intellectu* is always an *adequatio*; our representation of something, whether it is by way of a physical construct, or a diagram, or a logical or mathematical theory still has the aspect of our taking it to be a model – a property which its object doesn't share in the same respect if it is not identical with its model – and also is abstractive, in the sense given above: there are properties which the object of our modelling has and which we know it to have, which are relevant in the sense that we take them to be characteristic of the object, and are aspects of our inquiry and interest in the object, which are *not* represented in the model. And more than this: barring the claim to absolute knowledge about anything, we presume the object or state of affairs which we model to have properties we have not yet come to know about. The peculiar feature of successful models in science is that we may sometimes discover the properties which the object of our inquiry has, by the use of a model: it leads us, by the examination of its properties, to adduce certain properties of its object, which we then set out to discover by testing.

This would make it appear that the model, if a good one, continues to be *as rich* in properties as its object, if on continuous extension of inferences from its structure and characteristics, we have a preservation of reference to its object. But again, if this were true at the limit, the model would be identical with its object; and indeed, if we find properties of the model which we discover not to be exhibited in its object, then we delimit the relevance of the model. Mary Hesse describes this 'negative analogy' as a limit on models taken as intended factual descriptions thus:

A model is intended as a factual description if it exhibits a positive analogy and no negative analogy in all respects hitherto tested, and if it has surplus content which is in principle capable of test.[8]

Strangely, then, a model richer in properties than its object fails as a model because it is too rich, and 'negative analogy' can be discovered.

We can hope that it is rich enough for us to continue to use it fruitfully as an instrument of inquiry about its object; but barring identity of model and object, it cannot be as rich at the limit and yet remain a model. It must therefore be less rich in relevant respects than its object, which is precisely what we mean by characterizing it as abstractive. We may then reformulate (1), where $R(x)$ means the range or richness of relevant properties of x:

(2) $M(S, x, y) \& R(x) < R(y)$

which establishes the asymmetry of the modelling relation:

All this is by way of arguing for a representationalist account of models. But 'representation' then is taken in the broadest sense as any sort of mapping of structures on structures, or qualities on qualities. The essential feature of representation is reference, and it may be argued that not all reference is 'representational'. I would argue, perhaps perversely, that it is.

Unfortunately, 'representationalism' raises Lockean ghosts, and all the objections to the picture theory of language. The trouble, I think, is not with Locke or with pictures but with a simple – minded notion of picturing, taken as a paradigm of representing. Since, it is argued, scientific theories and models don't 'look like' what they are theories and models of, in the main (except for the trivial aids to the pictorial imagination which diagrams and Tinkertoy models provide), a representational account, as a copy theory, verges on nonsense. I would argue, on the contrary, that 'pictures', 'copies', even 'mirror images' are among the most extraordinarily complex cases of representation, and not the basic and simple ones. They are basic and simple only in one sense: they are the most familiar forms of representation that we know; our practical preoccupation with them has proceeded for several hundred thousand years, and so they form the basis of our common sense concepts of representation. But we know if we think about it, that that hoary example of 'direct' or 'exact' copy, the photographic picture, is an incredibly complex transformation of the properties (and only certain selected properties) of an object. Moreover, modern optics reveals the inordinately complex transformations which constitute the 'simple' mirror image: we only currently begin to understand what a mirror image is, though we have practiced practical reliance on such images for the whole lifetime of mankind. Perhaps ancient wonder and awe of the mirror image needs to be revived. So it is with all 'pictures' and with all

'picturing'. On analysis, then, 'simple' pictures turn out to be more complex transformations of their objects than most theories are.

I take models then to be pictures, allowed this complex sense: but pictures which refer, as some pictures do. The reference is always to something taken to be real, and outside the picture, or representation of it: no self-reference then, or no thing as model of itself. The sense in which a picture 'resembles' or 'looks like' its object may then range all the way from the familiar sense in which we may map contours on contours, from one to the other; but it may also be the sense in which the 'representative' of a nation represents the nation: again a mapping of his views, his votes, his behavior *qua* representative on the properties of that entity we take to be his nation. Insofar as the mapping fails, he fails to be a true representative.

The more radical thesis which I would propose – and this is the purport of my subtitle: 'Proposals for an Immodest Realism' – is that the modeling relation holds not only for the more complex mappings or representations we call scientific theories or theoretical models in science, but also for every descriptive utterance. Every sentence which may be characterized as descriptive is then a (small bore) model of what it is a sentence about: it is a representation of its reference – presumably some actual state of affairs. If this isn't bad enough, I would go on to claim that all utterances in a language are descriptive sentences, insofar as they are linguistic utterances at all – that is, are used to communicate in some community of speakers. Now this assuredly needs some explanation.

'The chair is on the table' may obviously be taken in a number of ways on different theories of language: it is certainly a picture, and I will not argue this. But it may be taken to express my belief, or to be symptomatic of a certain state of the speaker as a physical organism, or it may be taken as an inscription appearing four lines above this one. It is a model when it is taken by me, or by you as representing a state of affairs. But then what of 'Ouch'? What of imperative, emotive, performative, expressive utterances – the whole crazy zoo of distinctions of purportedly non-descriptive utterances? Take 'Ouch!' as the case in point. I would claim that it is true or false, that it refers, and that it refers plainly to a state of affairs, namely *my being in pain*. Similarly for imperatives: 'Open the door' refers, and refers plainly to a state of affairs, namely, to *my desire that you open the door*. Realism requires only that such desires, intentions, feelings, etc. are to be taken as states of affairs, and it seems trivially true that they are. A physical realist might argue that such

states of affairs are physical states of an organism; a more sanguine dualist might argue that they are 'psychological facts'; but that they exist, and that they are the reference of expressions in a language seems unarguable, though it is problematic how one ought to characterize them ontologically, or how one ought to relate them to other sorts of states of affairs. One may make a similar case for expressions whose purported reference is to 'formal' or 'mathematical' objects or relations, but then the argument doesn't concern reference, bur rather ontological frameworks. There are of course non-referring expressions, such as 'a' or 'is' or 'or', but they are not sentences, and not, properly speaking, linguistic utterances at all, but parts of them. (Trivially, we could say that 'a' refers to a, as its name, and I would argue that names refer, as elliptical expressions for " 'x' is the name of x", which is then a descriptive statement about usage, which is either true or false, and obviously context-dependent.)

But then it would appear that all linguistic utterances are models, representations of facts, or intended factually true descriptions. Conversely, that all models are one or another form of linguistic utterance, used to communicate an intended factually true description. Further, that the difference between individual, ad hoc occasion-sentences and standing sentence-complexes such as theories in science is not a difference in primary function – for all are equally representations – but a difference in the scope, and the ways in which this function is fulfilled.

If one takes this in a Braithwaitean way as saying that sentences are *theorunculae* (or to use a Yiddish diminutive which goes one step further: 'Theorenyunkelech') that's perfectly acceptable, for the point is to argue for a continuum among the meanest linguistic expressions and the grandest theoretical constructs of science. It is to do more, but even I hesitate to broach this now: that representation in art and in science have a common base, and a common function.

We begin modelling, therefore, with our first mimetic acts, and with our first use of language. And we continue modelling by way of what, on various grounds, have been distinguished as analogies, models, metaphors, hypotheses, and theories. The immodesty of the activity lies in this: if I am right, then more of our activity is truth-seeking than we know or acknowledge, generally. For models purport reference, and reference purports truth and falsity. There may be more analytic sense than poetry in the characterization of man as a truth-seeking animal.

NOTES

[1] Peter Achinstein 'Models, Analogies and Theories', *Philosophy of Science*, **31,** 4 (Oct. 1964), pp. 328–350; 'Theoretical Models', *British Journal for the Philosophy of Science,* **XVI,** 62 (1965), pp. 102–119; and in private discussion.
[2] 'Theoretical Models', *Op. Cit.*, p. 105.
[3] Marx Wartofsky, 'Models, Metaphysics and the Vagaries of Empiricism', this volume, pp. 24–39.
[4] Leo Apostel, 'Towards the Formal Study of Models in the Non-Formal Sciences', *The Concept and Role of the Model in Mathematics and Natural and Social Sciences*, Dordrecht, Holland: D. Reidel, 1961, p. 9.
[5] A. Rosenblueth and N. Wiener, *Philosophy of Science*, **12,** 4 (1945), p. 316 ff.
[6] *Ibid.*, p. 320.
[7] B. F. Skinner, *Verbal Behavior*, New York: Appleton-Century-Crofts, 1957, p. 124.
[8] Mary Hesse, *Forces and Fields*, p. 27.

REDUCTION, EXPLANATION AND ONTOLOGY

[1962]

In recent discussion of reduction in the sciences and its philosophical interpretation, the point has been made in different ways (by Carnap,[1] Nagel,[2] Quine,[3] Oppenheim and Putnam,[4] Bunge,[5] Popper,[6] and Hospers,[7] among others) that reductive explanation does not necessarily entail ontological reduction, that what is reductively explained is not necessarily explained away, and that reduction is not simply a case of elimination of the reference of reduced terms or reduced theories.

In these discussions, 'reduction' is used in various senses, and its career can be traced from the early positivistic elimination in the physicalism of Carnap and Neurath to the 'modest empiricism' of Scheffler[8] and the attack on meaning invariance by Feyerabend.[9]

A quarter of a century ago, Carnap introduced the notion of reduction in its purely syntactic form as a method of eliminating predicates by a substitution rule. But where a new term is introduced which has factual content, Carnap proposed the formalism of reduction pairs to modify the outright elimination which is possible with explicit definitions. This revision of the earlier positivistic reduction provides, then, for reducibility, but not "unrestricted possibility of elimination and retranslation".[10]

Nagel, in dissociating reductive explanation from ontological reduction, says:

The reduction of one science to a second – e.g. thermodynamics to statistical mechanics, or chemistry to contemporary physical theory – does not wipe out or transform into something insubstantial or 'merely apparent' the distinctions and types of behavior which the secondary discipline recognizes. Thus, if and when the detailed physical and chemical and physiological conditions for the occurrences of headaches are ascertained, headaches will not thereby be shown to be illusory. On the contrary . . . all that will have happened is that an explanation will have been found for the occurrence of headaches.[11]

In another place, Nagel defines the reduction of one science to another as:

the deduction of one set of empirically confirmable *statements* from another such set. However the issues of reduction are frequently discussed on the supposition that reduction is the derivation of the *properties* of one subject from the properties of another. . . .

The conception of reduction as the deduction of *properties* from other properties is poten-
tially misleading and generates spurious problems. . . . It is not *properties* but statements
which can be deduced.[12]

In a similar vein, Quine, in arguing for physicalism, offers the mentalist a
"less drastic version of elimination" which asks:

Is physicalism a repudiation of mental objects after all, or a theory of them? Does it
repudiate the mental state of pain and anger in favor of its physical concomitant, or
does it identify the mental state with a state of the physical organism (and so a state of
the physical organism with the mental state)?[13]

Oppenheim and Putnam, in discussing reduction,[14] declare as off limits
to their discussion what they call "epistemological theses," i.e. reduction
to some basic sensationalistic predicates or to the observable qualities of
physical things. They go on, however, to accept tacitly (and in one place
to suggest explicitly) an ontological context for the hierarchy of
reduction, at least in the sense of regarding a natural order of sciences as
plausible in terms of the "rough similarity" of the "well-known orderings
of things" to the reductive levels which they propose. In any case, the
explanatory reduction, on this 'levels' view, involves no ontological
elimination of subsumed levels.

   Most explicitly ontological are the views of Bunge, the name of whose
article – ('Do the Levels of Science Reflect the Levels of Being"[15]) – gives
the context of his interpretation. His answer is a qualified "no" and he
characterizes the relation as a "heteromorphism of being and knowing"
which leaves the question of an explicit ontological reduction as a matter
of scientific determination.

   In all of these views, explanation is interpreted as reduction. But the
senses of reduction, and the emphases are different. Reducibility may
mean, on the extreme eliminative view, explicit definability of terms, or it
may mean limited translatability and only partial specification of
meaning, or it may mean deducibility of sets of empirically confirmable
statements from other such sets; it may mean extensional identity of
reduced terms and reducing terms, or extensional isomorphism of
reduced theories and reducing theories, while intensional differences are
maintained, or it may mean constitutive (part-whole) or ontogenetic or
causal relation between elements in a science or between elements in
different sciences, or, on another view, of different levels of reality. But
in none of these cases is the reductive explanation synonymous with
radical ontological elimination and this is a point well worth making.

For the traditional objection to reductive explanation, more often than not, rests on the view that a reductive explanation commits one to ontological reduction of the 'nothing-but' sort. On this view, reductive explanation is objected to as an 'explaining away' which distorts or eliminates features of reality, and thus leaves them unaccounted for.

The error of the traditional view is that it confounds two dissimilar cases: one in which the reductive explanation explicitly entails an ontological reduction, and another in which there is no such claim. The traditional objection is meaningful only in the first case, but the decision as to whether the ontological reduction claim is legitimate is a scientific, not a philosophical one.

Examples of such explicit ontological reduction would be the historically settled theory-reductions of the Caloric theory of heat, or of Impetus theories of motion, or of the Phlogiston theory of burning, or of Devil theories of disease, and more problematic cases such as the interpretation of the null-results of the Michelson-Morley ether drift experiment, and nominalist attempts to eliminate classes.

But the traditional objection is meaningless if there is no explicit claim to ontological reduction and if reduction in the explanatory sense is itself meaningful. Now, it plainly is meaningful, as evidenced by the examples discussed previously. But although the view that reductive explanation does not necessarily entail ontological reduction *states* the issue in a correct way, and shows wherein the error of the traditional objection to reduction lies, the issue is not thereby resolved. For what is still at issue is the ontological status of reduced terms and reducing terms, of reduced theories and reducing theories, in a reductive explanation; in short, the question of the ontological reference of reductive explanations in the cases where there is no claim to ontological reduction.

There are three alternative approaches to this question which I wish to consider. The first two, I propose to call (1) *ontological evasion* and (2) *ontological postponement* (which fill out the triad of which *ontological commitment* is the third member). I will argue that both these views stop short of a consideration of ontological commitment in the *thick* sense although the second deals with ontological commitment in the *thin* sense. The third approach deals with ontological reference in two alternative ways, one of which is the *Identity* view and the second of which is the *Levels* view.

(1) The *evasion* is associated with the instrumentalist view of the nature of theories and of so-called theoretical entities. Its avoidance of the claim

to ontological reduction in explanation is by the radical device of disclaiming ontological reference in theories at the very outset. Obviously, since there *is* nothing to reduce, in this sense, there is nothing to be eliminated in a reductive explanation. And since reduction is a matter of the *deducibility of statements*, talk about elimination of entities or properties by reduction is meaningless, or spurious. This, then, is non-ontological reduction, (to which I would counterpose the alternative: non-reductive ontology). On this view then, theories have no denotation, and are neither true nor false. The criterion for theories is the instrumental one of adequacy, and the explanatory adequacy of a theory is that one can deduce from it empirically confirmable statements. Somewhere in the reduction, therefore, there must be empirical or operational reference. The asymmetry of reduction devolves then upon some 'basic' or 'fundamental' reduction-level, let us say observation predicates, or sense-data, or physical entities in space and time. But there's the rub. For if the theory is to be meaningful at all, and if it is going to be an adequate scientific theory, then the ontological neutrality of this view has been bought at a very high price, and the buyer has been cheated because neutrality has in fact not been guaranteed. For although one may say that it makes good sense to have *some* empirical reference at some level in the reduction, and one may add that empirical reference is after all not ontological reference, the argument has already taken a fatal turn: at the tail end of the evasion, one has run full tilt into the epistemological windmills. The evader is now required to present an adequate criterion for establishing the point at which so-called 'theoretical terms' leave off and at which the so-called 'observational terms' (or whatever reference one chooses) begin. The instrumentalist, neutral to the last in matters ontological, holds on to his observation-terms, his sense-data, his 'immediate givens' or their contemporary counter-parts, the 'quickly decidable sentences', to keep them from the ontic contagion. "I *assert* nothing," he cries, "I merely *adopt* one or another framework as it suits my needs and purposes. I don't even *believe* except in the sense that I can believe whatever it is adequate to believe, if I choose to." Now it may be true, as William James remarked, that the trail of the human serpent is over everything. But belief contexts involve at least the 'thin' ontology which is satisfied with a definition of reality as that which is *believed* to exist.

(2) The evasion turns out to be a way of backing into the second of the views to be considered, the ontological postponement. This view at

least sounds more promising. It does not deny ontological reference, but rather confines the issue to a choice among alternative reference schemes, which involves, in practical terms, a choice among alternative languages, each language carrying with it its peculiar ontological commitments, in terms of the values of its variables or the entities which are countenanced within the system. In what then is the postponement? Willy-nilly, whatever choice you make, you are ontologically committed, are you not? But the question of alternative ontological commitments is not, in fact, a descriptive question but a normative one. What is involved is not what ontology one is *in fact* committed to in choosing one or another language framework, but rather which framework one *ought* to choose. But the postponement view hedges at precisely this point. The appeal to some criterion of choice among languages is either denied or put off. The denial is the judgment that such an appeal is cognitively meaningless, and is, in Carnap's phrase, an 'external question'.

Carnap writes:

> To be real in the scientific sense means to be an element in the system. Hence this concept cannot meaningfully be applied to the system itself. Those who raise the question of the reality of the thing world itself have perhaps in mind not a theoretical question as their formulation seems to suggest, but rather a practical question, a matter of practical decision concerning the structure of our language. We have to make the choice whether or not to accept and use the forms of expression in the framework in question.[16]

But what are the 'practical grounds' of such a choice? If they are not theoretical, (that is, I take it, not cognitive), are they the 'practical' grounds that we choose in fact just what we do choose? Grover Maxwell, treating this question in an ardently Carnapian way, recently characterizes Carnap as "tantalizingly terse" on this issue.[17] That is graciously put.

Suppose, instead, we say with Quine that the quality of myth is relative. But are some myths better than others? Assuredly. On what grounds? That they better suit our needs and purposes, or that they are systemically more adequate, or they are simpler, or ontologically more economical, or because the reference is "more deeply rooted in our formative past".[18] This is at least promising. These are at least theoretical grounds, and can be investigated, difficult as this may turn out to be. But they can be meaningfully investigated only if these questions themselves are seen as ontologically relevant *outside* the ring of fire of 'choice of language frameworks'. Because in order to discuss the

criterion of choice of linguistic frameworks, it must also be discussed in some framework which we have chosen, on the basis of some criterion which is either 'practical' and non-cognitive (and we are caught in the denial again) or which itself needs to be cognitively justified (and we are caught in an infinitely regressive postponement again). Is it then impossible to ask why one framework should suit our needs and purposes better than another, or why ontological economy should be preferred to ontological proliferation? (Remember James' comment that "Profusion, not economy, may after all be reality's keynote".[19]) Is simplicity simply a *logical* question about the nature of theories, or is it an *empirical, and* an *ontological* question?

Pending the resolution, or at least the theoretical consideration of these questions, the *postponement* view, while it clarifies the question of ontological commitment, stops short of considering it *as an ontological question*. Instead, this view is, properly speaking, a consideration of metaontological questions, and is useful at least in this, that it refines the context for the consideration of ontological questions proper.

(3) Both the evasion and the postponement stop short, then, of a consideration of questions concerning the ontological reference (in the 'thick' sense) of theories, and therefore of the ontological significance of theory-reduction in the sciences. If ontological reference is purported, in the fairly straight-forward sense in which, for example, Popper asserts that every scientific theory makes the implicit claim to describe something real,[20] then there are two alternative approaches to the question of the ontological reference of reductive explanations, which I shall call the *Identity* view and the *Levels* view.

(a) The Identity view asserts that the reference of reduced terms (and theories) and reducing terms (and theories) is unique; i.e. that the reference of 'reduced x' and 'reducing x' is identical. The Levels view asserts that the reference of reduced terms (and theories) and reducing terms (and theories) is non-unique, i.e. each has its appropriate reference, in terms of ontological *levels* or *contexts*. In neither case, therefore, is there a claim to ontological reduction in an explanatory reduction.

If, on the Identity view, I assert that the reference is unique, then the differences between reduced terms and reducing terms are no more than different ways of talking about the same thing. Thus, for example, if in the reduced (secondary) language, I talk about genetic laws, and in the reducing (primary) language, I talk about the biophysical properties of

the DNA molecule, which explains these laws; or if I talk about the hardness and brittleness of a certain metal in one language, and in the other talk about the solid-state physical model of molecular and atomic structure which explains hardness and brittleness, I do not thereby eliminate either the genetic laws or the hardness and brittleness of the metal in favor of a more ultimate reality. Rather, the reference of both descriptions is unique. Although the 'one' may be said to explain the 'other', this is only a rational distinction, not a real distinction. For there aren't *two* things, genetic laws *and* DNA molecules, but only one thing, the DNA molecule in its mode of operation, which *is* the genetic law, or, to say the same thing differently, the genetic law, which *is* the mode of operation of the DNA molecule. So too, James, in the earliest statement of his theory of Emotion, writes:

The emotional brain processes not only resemble the ordinary sensorial brain processes, but in very truth *are* nothing but such processes variously combined.[21]

and later, in the *Principles of Psychology*, says that the feeling of the bodily reactions to some stimulus *is* the emotion.[22] He doesn't say it *causes* the emotion, or is the *condition* or the *correlate* for occurrences of the emotion, but *identifies* it with the emotion itself. Sellars echoes this view recently, in his argument against the sense-contents myth. For example: "The sensible qualities of things *really* are a dimension of neural activity."[23] So too, Joshua Lederberg: "When genetics has been fully reduced to its molecular foundation, it may continue to serve in the same relation as thermodynamics to mechanics."[24]

   (b) The *Levels* view, on the other hand, admits of two entities, one the reference of the reduced term or theory, and the second, the reference of the reducing term or theory. Here too, there is no claim to ontological reduction in an explanatory reduction. The reduced term refers, but its reference is explained as constituted by, or causally related to, or derived from (in the ontogenetic sense, not in the logical sense) the reference of the reducing term. In the ontology of the scientific world view, this means that tables can be 'reduced' to their constitutive cells, which in turn are constituted by molecular structures, which are constituted by atoms in a certain arrangement, etc., etc., (and probably *ad infinitum*). Further, that the person standing at the table can be similarly 'reduced' without losing one jot of his personality, nor with any risk that he will, upon atomic reduction, be ultimately indistinguishable from the table. Clearly, 'reduced' here does not mean 'deducible from', and I don't

think this common, ontologically referent use need cause any confusions, as long as we specify it, no more than saying 'the sun rises' would, if we know what we are doing when we say it. Now, the argument against reduction in this referent sense might run like this: "if one carbon atom is indistinguishable from another, and such atoms constitute in large part both person and table, doesn't the reduction ultimately deny the reality of the molar entities in favor of the micro-elements? Isn't micro-reduction ontologically directional? And therefore, doesn't the ontological status of any constituted entity settle like sediment to the bottom of the reduction hierarchy?" The Levels view would answer with the claim that each level in such a reduction scheme is as real as each other level; that the ontological directionality or asymmetry is one of part-whole relation, or of derivation, or of causal order; that insofar as the reductive explanation scheme is hierarchical in this sense, even if every last instantaneous characteristic of the person could be given in terms of a complete specification of instantaneous quantum states, and even if complementarity were to be resolved in a new departure in quantum theory which could permit instantaneous specification to any degree of precision of the canonical conjugates, still nothing would be dissolved or eliminated. The ontology remains hierarchical, the names of entities at each level *refer*, and refer to conjecturally and purportedly real things, states, events, processes, structures, functions, and whatever other characterization one decides to give to the objects of reference. (The saving grace is that such conjectures and such characterizations may be false, but there is good logical theory for taking care of such reference.[25]) The difference between such a view and the postponement view is that here the hierarchy of reduction is regarded as a real state of affairs, to which our descriptions in some sense correspond, even if "loosely and fluidly" (to use Bunge's characterization). If we claim that systemic adequacy is a criterion, we do so because we conjecturally regard the universe as systemically adequatable, and this conjecture itself is either true or false. If we use simplicity as a criterion, it is because we conjecture that nature works on a principle of least action. The test of the conjecture is not meaning analysis, and the test of the criterion is not that it suits our interests and purposes.

Rather, the meaning analysis *makes the conjecture testable*, and thereby gives it the status of a reality claim; and our interests and purposes *make the criterion meaningful*.[26] Being meaningful and testable are conditions for the discovery of scientific truth; these conditions don't

tell us what *is* true. Meaningfulness and testability give no greater truth value to one theory than to another. For what may turn out to *be true* about life on Mars, (within some specified time), or about sex-determination by genetic coding, is no less true now than it will be, merely because it is not yet clearly as meaningful or as testable as we expect it to become.

To summarize, then: both the Identity and the Levels views are similar in that both the reduced terms and the reducing terms are ontologically referent. Neither view evades the ontological question, nor postpones it, and neither makes a claim to ontological reduction. The reduction in both is a function of explanation, a relation among elements of our knowledge of the world, in which this knowledge is statable in terms of conjectures about reality which are true or false.

In terms of what kind of ontology each view tends toward, the Identity view tends to economy, the Levels view to proliferation. But the issue is not a conventional one, nor is the choice conventional, though we may choose one convention or another as our mode of conjecture.

Insofar as the Identity view limits its identity to extensional identity, it is an ontologically elegant view and tends in the direction of ontological economy. So for example, on the question of particle-wave duality, such a view would retain the ontological reference of the duality, but instead of saying that there are particles *and* there are waves, this view would say that there is that which under certain conditions behaves like a particle and under others behaves like a wave. This is a version of the classical double-aspect view. The 'that which' is a conjecture that an adequate model will be found (once the inheritance of classical wave and particle concepts is *aufgehoben*) which will *explain* this duality, rather than merely describe it. Once such a model is found – and this then is a program, rather than a proposal for a specific model – it will not eliminate waves and particles, but will find that they are extensionally identical with whatever more complexly conceived entity it is that explains them, and that therefore, the *meaning* of 'particle' and 'wave' has changed.

The Levels view tends to be ontologically proliferative, but systemically elegant. For if the Levels view has one overriding characteristic, it is its synechism, i.e. the continuity, the systemic unity and coherence of the hierarchy of reduction. Furthermore, it is marked for its tolerance for the introduction of new entities, and for differences among types of laws and principles of explanation appropriate to each

level. The explanation of one level by reduction to another involves ontological *direction* (micro-reduction) but not ontological *reduction*. Nor is the order of derivation, in the ontological sense, ontologically honorific. Thus, for example, humans are no more real than the 'lower' animals, nor are micro-entities more real than macro-entities. These leftovers of the older view of the normative status of the ascending or descending orders of reality, which runs through classical metaphysics and theology, is abandoned here. Ontological status is liberally distributed at every level.

What of the question of choice among these two alternatives, then? When ought we to adopt one view, when the other? When we discover, for example, that heat phenomena and light phenomena can both be accounted for in one theory, by one sort of entity; or when the laws of electricity and the laws of magnetism are seen to be statable in a single set of laws, or are seen to be derivable from one theory; or when and if the emotion or the sensation and its concomitant physiological state can be referred to a common theoretical model, e.g. of micro-states of the physiological organism, which will explain the fine-structure changes of both the physiological states of the organism, and the characteristic patterns of sense and feeling, or of some common *aistheisis* which subsumes them both; or when and if particles and waves will be discovered to be describable on an adequate 'wavicle' sort of model – then in cases like these the Identity view would seem to be a clear choice.

On the other hand, let us take a standard example like color. Even if so-called secondary qualities such as color turn out to be explainable as wave-lengths of electromagnetic radiation, *that colors appear* is a fact, and the *what-it-is-that-appears-as-a-color* is no less a factual state of affairs than the wavelength of light in the visible spectrum. It retains its ontological status at that level, though it is reductively explained. The description of *color-appearances* may be as physical as you like: it may involve the chemistry of retinal structures, of the rods and cones, and the physical interaction of wavelengths of light with such structures, and this would constitute a fact (though only *one* fact) of color-appearances. The phenomenological 'feel' of the color state may be a different kind of fact, related to this first one, just as in the first, two sorts of facts are related, i.e. retinal structures and wavelengths of light. In such a case, there would seem to be different facts, or collocations of facts at different levels, or in different contexts, and uniqueness of reference of

the Identity sort may impose on such a situation a constraint which is forced and artificial, and eliminative in the wrong sense. The Levels view would then seem most adequate to such a state of affairs.

The choice then is not to be made on philosophical grounds, but on scientific grounds: on the basis of what we *do* know about such states of affairs, in our theories, and on the basis of what we can fruitfully or interestingly conjecture. The philosopher's requirement here is the same as the scientist's, then: tolerance and forbearance in the uncommitted cases, and careful analysis in the committed ones; in short, the conditions for informed philosophical speculation and imaginative scientific theorizing.

NOTES

[1] R. Carnap, 'The Methodological Character of Theoretical Concepts,' *Minnesota Studies in the Philosophy of Science*, I, (Univ. of Minn. Press, 1956), pp. 74–5.

[2] E. Nagel, *The Structure of Science*, (Harcourt Brace & World, N.Y. 1961), Chapter 11.

[3] W. V. Quine, *Word and Object*, (Technology Press, 1960), p. 265.

[4] P. Oppenheim, and H. Putnam, 'Unity of Science as a Working Hypothesis,' *Minnesota Studies in the Philosophy of Science*, II, pp. 3–29.

[5] M. Bunge, *Metascientific Queries*, (Charles C. Thomas, Springfield, Ill., 1959), pp. 108–123.

[6] K. Popper, 'Three Views Concerning Human Knowledge,' *Contemporary British Philosophy*, 3rd Series, ed. H. D. Lewis, pp. 383–4.

[7] J. Hospers, 'What is Explanation?' *Journal of Philosophy*, (1946), reprinted in *Essays in Conceptual Analysis*, (Macmillan, 1960), pp. 94–119.

[8] I. Scheffler, 'Prospects of a Modest Empiricism,' *The Review of Metaphysics*, X, 3–4, (1957).

[9] P. Feyerabend, 'Reduction, Explanation and Empiricism', *Minnesota Studies*, III, 962, pp. 28–95.

[10] R. Carnap, 'Testability and Meaning,' reprinted in Feigl and Broadbeck, *Readings in Philosophy of Science*, New York: Appleton-Century-Crofts, 1953, p. 67.

[11] Nagel, *Op. Cit.*, p. 366.

[12] *Ibid.*, p. 364, 368.

[13] Quine, *Loc. Cit.*

[14] Oppenheim and Putnam, *Op. Cit.*

[15] Bunge, *Op. Cit.*

[16] R. Carnap, 'Empiricism, Semantics and Ontology,' reprinted in *Meaning and Necessity*, Chicago: U. of Chicago Press, 1956, enlarged edition. Second edition, p. 207.

[17] G. Maxwell, 'Theories, Frameworks and Ontology,' *Philosophy of Science*, 29, 2, (April 1962), p. 133.

[18] Quine, *Op. Cit.*, p. 234.

[19] W. James, *Pragmatism*, Lecture 5, New York: Meridian Edition, p. 126.

[20] K. Popper, *Op. Cit.*, p. 384.

[21] W. James, 'What is an Emotion?', *Mind*, 1884, reprinted in *Psychology Classics*, **I**, (Williams and Wilkins, Baltimore, Md., 1922), p. 13.

[22] W. James, *Psychology*, (1893), p. 373.

[23] W. Sellars, 'The Language of Theories,' *Current Issues in Philosophy of Science*, ed. Feigl & Maxwell, (Holt Rhinehart Winston, N.Y. 1961), p. 76.

[24] J. Lederberg, 'A View of Genetics,' *Science*, (1960), **13**, 3396, p. 269.

[25] Cf. Carnap, R., *Meaning and Necessity*, Second Edition, (Univ. of Chicago Press, 1956), pp. 36, 84.

[26] Cf. Nagel, *Op. Cit.*, pp. 363–64, on the claims concerning reducibility as 'temporally qualified questions.'

MODELS, METAPHYSICS AND THE VAGARIES
OF EMPIRICISM

[1965]

In what sense, in using a model, does one make or avoid metaphysical commitments? That is, in what sense does one commit oneself to claims concerning the existence of the entities, (relations, processes, events) which the model proposes? The pervasive use of models in both the didactic and theoretical aspects of science prompts this reconsideration of what is a classical problem in philosophy, concerning the nature of reality- or existence-claims, and inevitably then, of their warrant. More specifically, the question concerns scientific hypotheses, and whether, and in what sense such hypotheses make cognitive claims concerning what there is.

At the outset, I should plead guilty to a conflation of terms, the justification for which will be suggested as the paper progresses. 'Theories', 'hypotheses', 'models' and 'analogies' I take all to be species of a genus, and my thesis is best stated directly by characterizing this genus, as *representation* (though 'imaging' or 'mirroring' will do quite as well). Presumably, one should feel uncomfortable in thus reviving the ghost of Lockeian representationalism. The ghost's ancestry goes back, of course, to Plato's and Democritus' *eidolei* and to Aristotle's *mimesis*, and its metamorphic versatility is attested to by such a various past. At the outset, I want to claim no more concerning 'representation' than what the notorious plain man means by it. But what he means by it has two related senses: first, it connotes the relation of 'likeness' or 'resemblance' in which one thing is recognizably like another; and in addition may be taken to 'stand in' for it in some respect. Now in this second sense of 'standing for', there may be a weakening of the notion of 'resemblance' in favor of the functional emphasis on representation, as when one speaks of the representative of a nation, where there is no hard sense in which the representative 'resembles' the nation, though there is a sense in which he 'mirrors' its views, or its character or its aims, as a true representative should. I am aware of the arguments against so-called common-sense or naive representationalism, whether from illusion or hallucination, or from scientific 'two-tables' accounts, or from

sociological or perceptual arguments concerning the pervasive narcissism or anthropomorphism which transforms what is 'really other' into what is 'like me'. However deeply this relates to epistemological questions concerning metaphor and explanation, I want to bypass these here, and go directly to the ontological question in representationalism: namely the relation between what is taken as a representation, and the claim that what is represented exists. For plainly, not every representation is taken as veridical, and this veridicality is ordinarily what we mean when we say that what is represented does in fact exist. For what else could representation be veridical *of*, otherwise? Or *is* this in fact what we ordinarily mean? Obviously, a representation of Pegasus without wings would not be veridical; yet a proper representation of a winged Pegasus does not carry with it a commitment to Pegasus' existence. Whatever the Greeks may have thought, at various stages of demythization, we *know* that Pegasus does not and did not exist. The option of proposing imaginary entities as 'existing' in the conceptual space of human imagination, having properties there, and thus having veridical and non-veridical representations may be an artful dodge, but then proliferates meanings of 'exists'; and then we have a full zoo of 'existences', each animal marked with its proper habitation, as 'in mathematics', or 'in fiction', or 'in the world-view of the Hopi', or 'in Language-Framework *L*'. We may extensionalize the whole business, as One-Worlders, by discovering the intersection of various representations as their 'point of reference', but this in effect constructs an entity out of representations, (as a composite police-drawing of a suspect is constructed from the accounts of witnesses) leaving as yet unresolved the question whether the construct *refers*, i.e. whether such a purported composite or constructed entity exists. For this, we might go to the *intentions* of the compositor of such representations. If his source for the representation is the witness of observers, then presumably he (and they) construe the representation as of what they saw, and not of what they think the suspect should have looked like, or their imaginings of what he might have looked like, or what it might be useful to conceive him as having looked like, for one or another purpose. Only in this last case is there any tenuous tie to a purportedly existent suspect. In order to catch the real one, it is useful to represent him as he really is (or at least as he appears, if appearance is the instrumentality of his capture). But think how odd it would be to say that it is useful to think of a suspect as appearing in this or that way because such a useful fiction would most

successfully lead to his capture, and maintaining at the same time that we
cannot know what the suspect 'really' looked like. In short, such
ordinary recognitions as those of other people, or of tables and chairs
and the other objects of macroscopic perception are those in which
'purported looks' and 'real looks' are not distinguishable. Things are as
they appear, and naive realism flourishes on the identity. But naive
realism never makes the error of phenomenalism, in claiming that things
are their appearances, or that constructed resemblances are identical with
the things they are resemblances of. Nor does a naive realism concern
itself with the existence of appearances, or resemblances as such, for
these are transparent. But the Wittgensteinian charge that it is
philosophy which deviates and concerns itself hopelessly with the
existence of such odd entities as appearances, by departing from
common use, is a conceptually naive charge. Naive realism is
contemporary with naive animism, which concerns itself with the
existence or reality of images, whether as ikons, totems, dream-objects,
visions, or in some alternate representation in imitated likenesses in
painting or sculpture, or in mimetic dance or ritual, or in magic speech.
All this is pre-philosophical, and is as much a part of 'ordinary' use as
the paradigms to which Wittgensteinians are wont to appeal to set
philosophical pseudo-problems straight. But, we say, the age of witches
and demons is past, and contemporary image-making is disabused of any
existential intent. As in the arts, we are aware of the light which a
fictional portrayal throws on real human character, without yet imputing
'existence' to fictional characters, so too in science, we are aware of what
light models throw on the structure and character of nature, without
imputing existence to such hypothetical fictions. But models, like
characters in fiction, are not transparent as appearances are; they are
deliberate constructions, artifacts, and however much we consider them
as 'ways' in which something is understood to exist, as 'crutches for the
imagination', as 'computation devices' or 'inference machines', they are
themselves intermediate entities. Russell tried to obviate problems of
reference by distinguishing between inferred entities and logical
constructions, proposing that wherever possible we substitute the latter
for the former. So too, in psychology, during the decade of discussion on
intervening variables and hypothetical entities, the confusion persisted
that variables with values somehow eliminated the troublesome ontology
of believing something to exist. As Russell never made clear what, in
effect, logical constructions were, so too intervening variables remained

queer mathematical entities – in Hilgard's account, 'between-equations mediators' – without that 'surplus content' which characterized the existential claim of the hypothetical construct.

The impulse of ontological cautiousness is clear, and laudable. From a world populated with demons, witches, and vital forces, all constructs of the imagination mistaken for real entities, we have emerged upon the clear plain of empirically warranted assertions. However queer the new entities which rational scientific theory has introduced, they are anchored somehow in that careful and testable objectivity of experiment and observation. All that is needed is a philosophically viable account of the difference between demons and mu-mesons. Contemporary empiricism set itself this task, in the formal analysis of the conditions of empirical significance, of confirmation and falsification, of the truth conditions of propositions. Here, a logic without ontology proceeded to analyze the discourse of science, and to reconstruct the relations which existed linguistically between the theoretical and the observational terms in scientific theories and laws. And here too, Popper placed the demarcation question, seeking the formula which would cleanly distinguish scientific from non-scientific assertions. I have argued in another paper[1] that the reductionist program of early positivism, and its modifications in contemporary logical empiricism have failed, on internal grounds alone if on no other; and that the demarcation criterion breaks down under analysis as well. And others have argued similarly, challenging the sharp distinction between theoretical and observational terms, and the very viability of some unambiguous reduction base. In some quarters this is greeted as a return to good old metaphysics, to existential encounter with truth and being, to the robust speculation of the greater ages of philosophy. Perhaps it is time for a post-prohibition binge. But sober morning-after thoughts for serious philosophers will not dissolve in inebriation. Is empiricism really in such desperate straits? Is the old game of checking with the facts to be put so unceremoniously aside, or relativized to contexts in quite the way that the permanent-revolution boys suggest? Is empirical objectivity to be replaced by obscurantist intuitionism, or by a new, more rigorous rationalism or coherence theory? These are programmatic questions, and although it is fun to conceive of philosophy as ideological battle, it clarifies little.

I propose therefore to raise the question of what can strengthen and give blood to an anemic empiricism, by way of a detour; that is, in

considering the metaphysical commitments a conscientious empiricist may make, particularly in his use of models. I would claim further that scientific empiricists ought to make such commitments, and I will suggest why later.

But what exactly does it come to to say that an empiricist makes metaphysical commitments (or that anyone does for that matter) and why should it be problematic to assert this? Is the claim that theoretically postulated entities *exist* such an expensive one, either in science or philosophy, that one can afford it only seldom? Or is the fear of reverting to unscientific and uncritical hypostatization so great? Or perhaps it is the common-sense or ordinary-language-oriented intuition which balks at using 'exists' in the deviant ways which new scientific models press on us? Is it primarily a matter of linguistic or conceptual comfort and discomfort? Clearly, the central questions are two: (1) In what sense, in using models, does the scientist make any existential claims? (2) Are such claims necessary or dispensable? One qualification here: plainly, an existential claim as such is not 'metaphysical'. A metaphysical commitment, if one takes this in the serious sense which the history of philosophy would suggest, is a systematic one which links up a claim that this or that exists with a conceptually critical and rigorously systematic framework within which this claim is consistent with more fundamental and general postulates, i.e. with some categorial scheme, and some formal structure of relations among things – in short, with a metaphysics. The reason that the question of metaphysical commitment comes up with respect to the existence claims of models is that models are systematic, and that a minimum or sufficient condition for something to be a model is that it postulate in some systematic way not merely entities, but feasible relations among them. In this sense, every theory is a model, as every metaphysics is a model. But so too are the more modest systematic constructs like subway maps and cookbook recipes. The more difficult conception that language itself is a model (apart from the question of models *of* language) forces the representationalist thesis to its hardest test. But I will avoid this thorny question here.

The compatibility of alternative models (as e.g. of waves and particles), or of models of different domains (as of physics and psychology) raises the additional question regarding the scope of the model, its degree of *adhoc*ness, and thus the quality and generality of its existence claims.

Preliminary analysis therefore demands a review of the uses of models,

and I will range them hierarchically, from the weakest to the strongest existential claims. In the weakest case, there appears to be no serious claim that anything exists; no cognitive claim therefore, but only an invitation to imagine or conceive of things in a certain way, for what it's worth. In the strongest case, the model is taken as an "intended factually true description" (in Kemeny's phrase), commanding (or at least commending) belief and assent. The list runs something like this:

(1) *Ad Hoc Analogies*: These are the familiar qualitative models which appeal most directly to striking but isolated resemblances between a *representandum* and a *representans* (borrowing here the current Latinate barbarisms, with due apologies to good taste). Thus, if we are told that stellar systems in an expanding universe are like raisins in a rising lump of dough, or that the body politic is like an organism, with its head, eyes, ears, arms, legs, stomach and heart, or that electron orbits in the atom are like shells, the analogy presumably does no more than to aid the formation of a conceptual picture. Its uses may be said to be didactic, rather than scientific in any full-fledged sense. Thus too, mechanical demonstration models, with their springs and wires – what I would call Tinkertoy models – call attention to physical relations and qualities which in no sense are purported to be literally factually true. One may draw a distinction here between formal and qualitative analogies, as Achinstein has done, and grant that the physical interpretations given to various expressions in the analogy have an independent and contributory role apart from structural or logical analogues. But the *ad hoc* character of such analogies tends to defuse any ontological bomb that may be hidden in them. Nothing more is claimed than that certain features which we recognize or are familiar with in the analogue are usefully like certain features in what it is an analogue of, for conceptual or picturing purposes. That no more than this is claimed is clear from the acceptance of the negative analogy in such cases, so that Mary Hesse marks off such limited claims by considering them conjunctions of positive and negative analogy (or disanalogy). In short, we know what we are doing when we use them, and we know that the analogy is factually limited, and false if extended.

(2) A different, but no stronger claim is that made in the formal representation of a limited domain of experimental facts. The impulse here may be economy of representation, and this is the classic model of Duhemian 'laws' or of Mach's 'economy of thought'. At the lowest level, this is no more than data-ordering, or shorthand-description, and the

only claim that accompanies this is that the experimental facts may be ordered in this way; anything beyond this, as an interpretation of the ordering, goes beyond such description, and thus beyond the limits of this 'model'. And yet there is hidden here a stronger claim: for such models – of which so-called mathematical models in the social sciences and in psychology are paradigms – have a more systematic claim: namely, that such models hopefully exhibit the law-like properties of the data, that they are in effect simulations of large-scale and universal behavior of the variables, and represent, therefore, inductive generalizations from a sample. They then may also be taken as hypotheses. But there is only a limited ontology in this, for the 'entities' of the model are data, and the interpretation of these data lies outside the model. One of the chief contributors to mathematical models in the social sciences, Simon, writes in his *Models of Man*, "In this paper we have constructed a mathematical model that appears to translate with tolerable accuracy certain propositions asserted by Homans to hold for the behavior in human groups", and in another place: "The concepts to be defined all refer to a model – a system of equations, and not to the 'real' world the model purports to describe". Sternberg, in his contribution to a volume on mathematical psychology, writing about Stochastic Learning Theory, says "Experimental events, with assumptions about their equivalence and complementarity determine a set of model events and thereby give rise to four important classes of models . . ." and elsewhere, "The crux of a model is its description of response-probability changes from trial to trial. . . ." In these cases, there is a tacit assumption that data gives us a handle on the world and the way it is. But the model's reference is not to an already interpreted 'world' but rather to the direct and intermediate phenomena which constitute the experimental events. The model, in short, deals with a universe of laboratory entities, to be accounted for and 'pictured' mathematically. The Galilean-Pythagorean assumption that mathematics is the language in which the book of Nature is written lies close beneath the surface of such mathematical model-construction, but it is generally suppressed, or qualified (e.g. by quotes around 'reality') and most easily passes over into the instrumentalism which we will consider next.

(3) Models are computational devices, or inference machines. Their validity lies in their structural or formal isomorphism with the set of descriptive statements and their relations, in some empirical science.

They serve therefore to draw the consequences of theories, and thus to serve an auxiliary function in leading theories to the test. The function of a model is therefore to generate predictions which are testable. On some such view, the burden of commitment is passed on to the theory of which some computational model may be constructed. The postulates of the theory may make existential claims, therefore, but the model serves merely to channel these to some confrontation with experimentally testable consequences. This is a reasonable construal of what it is to be a model of a theory, and the model-theory distinction serves a relatively useful function then. But in more decisively instrumentalist views, the theory itself serves this same function, being itself no more than an instrument for coherent organization and testing, and the question remains – *of what*? The reference beyond such theory-model 'instruments' remains forever delayed; or it is defined in terms of practical purposes, decisions concerning which lie outside the theory, but are vaguely defined as 'successful prediction' or 'control of the environment'.

(4) 'As If' constructions: On this view, theoretical or hypothetical models are seen primarily in their psychological function, or in their methodological utility in scientific theory-construction. They serve the heuristic function of rendering a complex theoretical domain intelligible, or of suggesting directions of thought. In this, they are like *ad hoc* analogies, but make a stronger operational claim, tending to be more systematic and of wider scope. They represent another instrumentalist emphasis, on the utility of such "cognitive maps" (to use Tolman's phrase, though he was describing rats), or "plans" (to use Miller, Pribram and Galanter's name, which is OK for humans too). But here too, there is no explicit existence claim, beyond the abstemious one James might have characterized as "What it is better to believe". Now it might be better to believe the truth, but the cautionary virtue of 'as-if' constructions is that there is no claim to truth, unless it be the suspect one that what it is better to believe is true by virtue of that vague criterion. Here, however, one may introduce an intermediary interpretation of 'as-if' models, which borders or teeters on the edge of a cognitive claim. It is that the 'as-if' model is an 'abstract model', which although it has as yet no filling out beyond the limited domain in which it is operative, suggests that it may, with further scientific advance, become 'concretized'. I am thinking in particular of the currently much-discussed Chomsky model in linguistics, which purports a hypothetical 'device' which effects the transformations which a user of a language would have

to make, to generate grammatical sentences. Wiener, discussing such 'black-box' explanations (in his *God and Golem, Inc.*, and elsewhere) draws a distinction between 'pictorial' and 'operative' images which bears on this. What goes on in the 'black-box' (say, Chomsky's 'mind') is forever sealed off from 'view', in the ordinary sense of revealing to us the 'picture' of what goes on inside. Far from taking this as a constraint, Wiener (and I think Chomsky too) proposes that the *operative* image is the one that gives us a deeper insight into what is imaged, a more adequate isomorphism than the merely homomorphic 'pictorial' image, which represents essentially superficial and even accidental features. The question, "Is what is in the 'black box' really like that" becomes redundant on this view, since all that we may ultimately require of a model is that it gives us an adequate functional account of transformations. If we take this one step further, however, we arrive at the stronger claim that the black-box, when eventually 'opened', will reveal just the sort of things we hypothesized. What this comes to, in effect, is that physiology will redeem psychology, or more radically yet, that it will ultimately have to, since the psychological or psycholinguistic model is a true representation of the psychological or the psycholinguistic facts. Such a claim goes over from the instrumentalist position of entertaining a hypothetical entity, or a hypothetical structure, *per suppositionem*, to the quasi-realist position which says, "Things are something like this, in point of fact".

(5) The models which make tentative cognitive claims are next in respect of the strength of such claims. These are approximate representations taken to be 'true' of a limited or abstractive domain of facts, and not known to be false upon extension. Here, Mary Hesse talks of "positive plus neutral analogy", and on such a view, the model is subject to cumulative modification and refinement, as the facts come in, approaching the 'true model' asymptotically. The question arises whether in an approximative representation, one is absolved from metaphysical commitments. One may always say that, being approximative, the model is always false if taken literally. Duhem dealt with this at length, in *Aim and Structure of Physical Theory*, in discussing the infinity of mutually exclusive mathematical formulations which would all fall within any approximation in the data. This raises the problem of so called 'ideal models', and these may be taken in two ways: either we are Platonists, and take the ideal model as the representation of what's 'really' real, writing off our approximations to the limits of a debased

empirical knowledge; or we take the model as an 'ideal fiction', required of us by way of the economies of ordering. We can manage only a finite abstraction of the infinitely textured world, and while God's eye may be on the sparrow, ours is only on its velocity, its mass, and its genus. Hegel's requirement that truth is only of the whole, and that all approximations are false, because approximate, restates this essentialism in its most absolute form. But the other side of this coin is equally classical fallibilism, in Peirce's or another version. The claims of the approximative model are ambiguous, somehow: for to claim that nothing exists to which the model is an approximation is to make logical hash out of the concept of approximation; yet, to acknowledge that a model is approximative seems to withold its existential claim in view of some superior knowledge which goes beyond approximation. One can know one is close to home, if one knows where home is, or what it is like to have arrived there, but not otherwise. The next stronger version of the tentative cognitive claim is to grant fallibilism its due, but to make universal and not approximative claims, as exploratory probes as to what is. This gets tied up, in philosophy of science, with considerations of the status of statistical and probabilistic assertions. But the useful ones tell us exactly what we want to know, and are falsifiable in their own terms; they bear less on the issue than they are usually made out to, and are, in their own manner, interpretable as universal claims. The universal claim is the typical hypothetical construct, taken seriously. It is tentative in the trivial sense that it is hypothetical. The assertion is put forth to be tested, and not simply accepted on faith, or because it suits one's taste. It has therefore a certain initial plausibility, either on evidential grounds, or in terms of its compatibility with other theories or models, or because it is metaphysically more nearly acceptable than its alternatives. But it is not approximative in the sense mentioned earlier; rather, it makes an all or nothing claim: it is either true as it stands, or it is false; and if false, then not capable of being accomodated in some higher synthesis, but rather another instructive error with which to pave the road of scientific advance.

Models taken in this sense are then what Popper calls "Conjectures", and purport to be descriptions of some factual or real state of affairs. Their condition of significance (though not of 'meaning' in any linguistic sense) is their refutability, or falsifiability.

(6) One still stronger claim, which eschews tentativeness, and goes to the limit of any rational belief, is that the model is a true description of a state of affairs — in Kemeny's phrase, "an intended factually true

description''. It would be tempting to conflate this with the 'conjecture' above; but the conjecture, while put forth as a universal claim, of unlimited scope and without approximation, derives its force from the skepticism which accompanies it. It is made in all good faith, but it is made for martyrdom. It is, like a defendant in a French court, provisionally guilty until proven innocent, and total innocence is forever beyond its attainment, by virtue of some Popperian Original Sin. The absolute cognitive claim — though it sounds reprehensible to the philosophic ear — is the commonest of all our ordinary claims. The degree of conviction which accompanies it is such that any challenges to it are recast so as to accomodate and maintain the belief. Nor is this the attitude of the marginal fanatic or dogmatist. It is common-sense realism, and scientific realism in those corners of science which have attained to the status of common sense. That there are tables and chairs, and food and love, and night and day is not atheoretical knowledge, though strictly speaking we may call it prescientific. Nor does empirical evidence shake us in these convictions, no more than the witness of our eyes leads us to believe that the magician, in pulling a rabbit out of a plainly empty hat, had in fact produced something out of nothing. Well, parts of science are like this. There is no real sense of tentativeness in the view that organisms are composed of cells, and that genetic inheritance has something to do with intracellular structures, however these are finally doped out. It may be argued that these are not 'models' anymore, for the model and what it is a model of are identical, in these limits. We can see cells, and watch mitotic division; likewise, we can carefully construe what chemical transformations are taking place and what structures are involved from the expert evidence of a world full of laboratories. Nor is it really a matter of conjecture that there are such things as molecules, though there is less comfort with outright claims about atoms, since the models have changed so rapidly since the turn of the century. At any rate, enough models in science have this characteristic of absolute cognitive claim, and direct existential claim so as not to make this case seem queer or marginal. Just as Eddington's 'two-tables' argument was unconvincing, if it was taken to conclude that there really aren't such things as the ordinary tables of our acquaintance, so too arguments about cells, or about the position of bodies in the solar system, etc. would be unconvincing, if they were taken to conclude that what we know just isn't so. At most, such arguments would suggest that there are intersecting alternative frameworks within which these things

may be conceived; but we would expect that on the whole, a comfortable relation would exist among these alternatives.

It is in the last case, then, that the strongest existential claim is made, and to which Tarski's semantic criterion of truth applies most unambiguously: i.e. that a statement is true if and only if the state of affairs it purports to describe exists. It may sound unordinary or extraordinary to say that the utterance 'It is raining now' *represents* a state of affairs, in any 'model' sense, and also unidiomatic and 'deviant' to say that *raining now* is a state of affairs which exists (or doesn't); but there is very little odd about it except the abstract terminology 'state of affairs' and 'exists', and there is really no trouble with this. The fear, of course, is that 'state of affairs' will be taken as an abstract entity, and said to 'exist' in some odd way. Then 'production' of a 'state of affairs' may be criticized as being parasitic on 'production of rain', and may conceivably lead to ontological monstrosities. But such wildly bad metaphysics has hardly ever been done, except in bowdlerized examples given by early positivists who didn't bother reading metaphysicians with care, or who chose choicely bad metaphysics as their examples or their paradigms of metaphysics in general. Philipp Frank was closer to the truth when he spoke of metaphysical construals of science as retranslations of scientific concepts into the language of common sense, but wrong when he saw this as merely regressive. Rather, I would say, models do mediate between science as theoretical language and common sense, but in the essential way which empiricism has always taken for its own: namely, the interpretation of abstract and relatively theoretical terms and structures in terms of relatively atheoretical, or relatively uninterpreted contexts. Empiricism chose 'sense-perception', or 'sense-data', or 'experience' for its reduction base, the model providing an interpretation of the theory in these terms, as a semantic (and not merely formal or syntactic) construction. But the vagaries of empiricism have led it to the impasse of an undefinable reduction base, or at best, one so queasily defined as to be no less abstract than that which it is to reduce. The concepts 'physical language', 'basic sentence', 'quickly decidable sentence' et al. are no less theoretically infected than are 'electron', or 'phase-space' (or 'sense datum', for that matter). But in both cases, something is purported to exist, if in fact a claim to truth is being made, whether in the 'theoretical' or the 'empirical' or 'reducing' or 'observational' language. The difference is that one language has been developed, at a given stage, as the language of thought – i.e. as a

language which countenances abstract, conceptual entities, and which deals with their relations in terms of formal and symbolic manipulation – and the other has been developed as the language of experiment, and of practice, which deals with the world of our perceptual-motor activity. That these are not really two languages, but two emphases within one complex language is what makes a rational empiricism possible, and permits us to relate theory to practice. Historically, or sociologically, it is the distinction between the language of the streets, of the trades, of the shop on the one hand, and the language of the religious ritual, of the reflective inquiry, of the court and the monastery on the other. Where inquiry grows effete, the learned language becomes distinct and its ties to the common language are broken. The Mandarin dialects of science are not the ones which the 'common people' don't understand, but the ones which are not related to the 'shop talk' of the laboratory, i.e. which are not related to testing and experiment. Existential claims and metaphysical commitments are likewise empty, and evitable in science where their coupling with our perceptual-motor experience and its expression in the language of this experience is weak or non-existent. This was the emphasis which operationism sought to make clear; the emphasis was sound; but its theoretical formulation in operationism was not, as Agassi pointed out recently. But the language of experience is not brute, nor atomistic in the old verificationist way, and here was the error of that militant empiricism. The language of experience – i.e. everyday language – is instead a highly refined and polished instrumentality, and a good one not on vague or *ad hoc* grounds of this or that particular and local need or interest. Rather, it is a good instrumentality, because it is the most universal representation of a common experience, and so primordial to what we mean by *understanding* or *making sense* that we don't make sense outside it, and we may be said to 'make' sense out of it. I have argued elsewhere that what accounts for the heuristic force of metaphysics in science is that metaphysics, when it makes sense – that is, when it is not bad metaphysics – does so because it is the most rigorous and systematic inquiry into this *universal* representation of common experience, and expresses its fundamental structures in the most felicitous way. The secret of why the pre-Socratics strike us as modern, or why we may still significantly speak of Platonism, or Aristotelianism, or realism, nominalism and conceptualism is that the issues go to the base of our common understanding, and are not the esoterica of a privileged

and idle intelligentsia, as they are sometimes made out to be. Let me put it in a way which brings us back to the question of models, therefore: the model is a more or less conscious metaphysical claim, not in being grandly philosophical in its pretensions, but in trying to get things into intelligible form for that deep-rooted common-understanding which is to science as the earth is to Antaeus.

The thesis would make little sense if this common-understanding were not *au fond* empirical in the best sense; but its empiricism would make little sense, if the warrant for existence claims were non-empirical. This is really what is at issue in the interpretation of models. The alternative interpretations, formalist-conventionalist, pragmatic-instrumentalist, or realist, are critically different only with respect to interpretations of the relations of empirical knowledge to truth. No news, this, certainly. But consider now the varying degrees of existence-claims we have examined. In every one, from the weakest to the strongest, from the most *ad hoc* and instrumental to the most blatantly absolute cognitive claim, the model purports to give us some way of understanding what it is a model of, either by qualitative analogy, or isomorphism, or in terms of computation or inference, or as a hypothetical construct. This we might in general label the heuristic content of the model. But Kant already pointed out in discussing regulative ideas, that heuristic force is no argument for cognitive content. However, if we look to simply empirical warrant for existence claims, this fails us too. Feyerabend has recently pointed out what sort of dogmatic metaphysics such an empiricism can lead to. The empiricist may, in effect, make no more than a methodological claim for his model: its entities, he will say, (whether 'segments of behavior', or 'operations', or 'protocol sentences') are meant to do no more than mirror his actual behavior as a scientist; let the metaphysician say more if he must. But this easily passes over into the ontological claim that what there *is* are 'segments of behavior', 'operations', and what it is that protocol sentences report: the explanatory and methodological reduction becomes an ontological reduction, and (to continue an already-begun play on words), ontology recapitulates methodology. Such an operationalist 'realism' for example, is what, in its worst interpretations, the reduction program of empiricism has become confused with: it is the simplest and sparest and also the least plausible of all ontologies: what there is is what is revealed in empirical laws inductively arrived at and experimentally tested. The reasons existentialism may have some force among scientists and non-scientists

alike, is that such an incredible reductionist ontology flies in the face of the common understanding, and conflates the scientific reconstruction of the world picture with our more complex experience of a world, leaving the gap to be filled by faith (whether 'animal' as Santayana has it, or something else), or by some 'immediacy' which is suprarational or arational. The dilemma this presents is the following one: Either (a) existential claims (or 'reality-claims') are empirically unwarranted or (b) if they are empirically warranted, in the reductivist version, the world is completely unlike what the common understanding takes it to be. The sub-dilemma for the empirically warranted version is that (b)' either 'reality' changes as science changes – in a realist's nightmare in which reality constantly transforms itself into exactly the image which science arrives at, at any time (and hence science turns out to be analytic and infallible) or (b)" our objective knowledge of what exists is so irremediably relativized, that 'truth' always comes labelled with a date and a location for which it alone may be said to hold.

I think the dilemma can be avoided only on the condition that some semantic criterion for the truth and falsity of models is adopted, and if one elaborates the sense of what may be said to 'exist' for a model to represent. Thus, even at the weakest, the feature which the *ad hoc* analogy singles out either does or does not correspond to its analogue in qualities which both may be said to have. That this is scientifically trivial is clear from the fact that any two things may be shown to be related in *some* way, and so the cognitive content of such a model is relatively empty, (though not analytic, for one may be mistaken even in the limited analogy). Anything beyond this, as Agassi has pointed out, is one or another form of generalization: whether a mathematical model, or a hypothetical construct. The relations between equations, in the model, certainly mirror *something*, and the elaboration of what may be said to exist would include here the relations which values of variables have to each other in the model, as representations of some such relation, or a transformation of it, among the 'data' (the erstwhile and circumscribed 'entities' of the purely descriptive mathematical model). The case is clear for the 'As if' models, and for those which make more or less explicit cognitive claims. If the argument runs: "I don't really take the entities in the model to exist, but it is useful to think of them that way in order to pursue a point, or to conceive tests for a theory with cognitive claims which the model somehow images", then I would raise the question as to what makes it useful to think of it that way at all, if there were not some

sense in which the model mirrored some aspect of what it is taken to be a model of. In short, the existence claim of a model may be limited in scope or applicability – i.e. in systematicity or generality, therefore – but to deny it such a claim makes a mystery of its significance altogether.

The dilemma's first horn, to make reality or existence claims warranted on non-empirical grounds, is not acceptable to a philosophy which takes scientific empiricism seriously. Neither is the second horn acceptable, in either of its forms. The alternative is to reformulate the grounds of existence claims, taking them back to their source and function in the common understanding. A formalist reconstruction of science may get along without them. I don't believe the scientist can, if he sees in science an inquiry into truth. From what I have said about the relation of metaphysics to the common understanding, the alternative would seem to be to relate metaphysics to a critical empiricism (or empiricism to a critical metaphysics), where the relation of heuristic to cognitive functions of scientific thought becomes significant and not accidental.

NOTE

[1] Marx W. Wartofsky, 'Metaphysics as Heuristic for Science', this volume, pp. 40–89.

METAPHYSICS AS HEURISTIC FOR SCIENCE*

[1965]

Now that the anti-metaphysical crusade of classical positivism has spent its force, and has been fragmented into the qualified and revisionist versions of logical empiricism, there is evidence of a cautious rediscovery of the relevance of metaphysics to science, within some recent discussion in philosophy and history of science. I say 'rediscovery' because the thesis is certainly not new, and some hardy souls within philosophy and history of science have held it all along in one or another version, even in the heyday of verificationism and reductionism. But what appears in present discussion is not radical enough. Rather, I would characterize it not simply as cautious, but as an attempt at piecemeal reconstruction *within* the framework of logical empiricism; or else simply as an emasculated descriptivist thesis about the history of science (simply repeating what every serious student of the subject knows: namely that metaphysics has always been relevant to science in paradigmatic historical instances).

The views concerning the relevance of metaphysics to science which I wish to examine are those of Karl Popper, Joseph Agassi, and Thomas Kuhn. I will argue that none of them go far enough. I will not simply argue *that* metaphysics was, historically, and continues to be, a heuristic for scientific research and theory-formation; but rather that it is inescapably so, both for the scientist who recognizes it and for the one who does not. Further, I want to show *why* it is a heuristic and how it functions. In short, I will argue that metaphysics represents the most general method of articulating, in critical and systematic form, the alternative conceptual frameworks within which theoretical understanding becomes possible. Further, my thesis is that the heuristic force of metaphysics lies in its closeness to our primary modes of understanding and explaining: by means of the *story*, the re-enactment of nature in dramatic form. I will argue that metaphysics represents these primary structures of the understanding in a fully articulate (and hence, criticizable) form in terms of three basic features: *reference, structure*, and *abstraction*; and that these, in turn, constitute the conditions of scientific theory formation.

Mine is therefore an explanatory thesis about scientific concept-forma-
tion, and not simply a description of the facts, nor a thesis in the history
of ideas. It bears on the question as to whether metaphysics is cognitive
(and either true or false, in some sense), and on the distinction between
good and bad metaphysics.

This paper is divided into four sections:

(1) I will examine the sources *within* logical empiricism and its self-
critique, of the recent 'rediscovery' of metaphysics (a) in the failure of the
early positivist reduction program intended to eliminate metaphysics both
from science and from philosophy, and (b) in the liberalization and rela-
tivization of the criteria of meaning and of empirical significance in
recent logical empiricism.

(2) I will briefly sketch the classical alternative positions on the rele-
vance of metaphysics to science.

(3) I will examine the views of Popper, Agassi, and Kuhn as important
recent[1] counterproposals to the positivist thesis.

(4) Finally, I hope to explain *why* metaphysics is a heuristic for science,
to examine the notion of 'heuristic', and to show in what ways Popper,
Agassi, and Kuhn fall short of accounting for the relevance of metaphysics
to science, even though they all agree that it is relevant.

I. THE FAILURE OF THE POSITIVIST AND LOGICAL EMPIRICIST
PROGRAMS FOR THE ELIMINATION OF METAPHYSICS

An important aspect of the advent of contemporary philosophy of science
is the anti-metaphysical crusade of modern positivism. A central enter-
prise of the pioneering work of Mach, and of his heirs (Schlick, Carnap,
Neurath, Reichenbach, Von Mises, Hempel, and Feigl *inter alia*) was the
elimination of metaphysics, as representing what was conceived to be
obscurantist, empirically meaningless, and therefore anti-scientific (or at
best non-scientific). From Mach's earliest formulations of the principles
of economy of thought and of the unity of science, through the develop-
ment of logical positivism and its revision in logical empiricism, a two-
pronged attack was undertaken to eliminate metaphysics: (a) by means of
a theory of meaning and reference which presumed to show that meta-
physics was meaningless, or literally non-sensical, and (b) by means of a
reduction of classical ontological alternatives to alternatives among

language frameworks, thus eliminating metaphysics by substituting prag-
matic criteria of expediency as the grounds of choice among such lin-
guistic frameworks.

Both of these attacks have foundered (one should add, largely on the
basis of internal criticism by logical empiricists and analytic philosophers
themselves). The demise of verificationist theories of meaning is too well
known to require a summary here. The modifications – replacing verifica-
tion by confirmation, by confirmability-in-principle, with the notions of
reduction-sentences, quickly-decidable sentences, acceptability – so dulled
the edge of the anti-metaphysical knife, that it is not at all clear what
surgery it is capable of performing at all. A recent critic writes,

> The neat and clean-cut conceptions of cognitive significance and of analyticity
> which were held in the early days of the Vienna Circle have ... been gradually
> refined and liberalized to such an extent that it appears quite doubtful whether
> the basic tenets of positivism and empiricism can be formulated in a clear and
> precise way.[2]

The same critic writes (of the notion of "empirical significance as testabil-
ity by experiential findings") that insofar as it is applicable to theoretical
systems, under the now-liberalized procedures introduced by internal
criticism, one could encompass even "axiomatized metaphysics of Being
and Essence" as empirically significant. It is enough to add that the critic
here is C. G. Hempel.

The attempted paraphrase of ontology, by a shift from ontological to
linguistic frameworks, also suffers on close inspection. In a recent evalua-
tion of Carnap's contribution in this regard, Y. Bar-Hillel holds that there
is a "decisive advantage ... in the transition from 'ontological' disputes
to methodological controversies".[3] The 'advantage' consists in reinter-
preting *assertions* as *suggestions*. So (using Bar-Hillel's own example), the
alternative 'assertions',

(1)   Numbers are classes of things
(2)   Numbers belong to a special primitive kind of objects

are interpreted in the formal, 'suggestion' mode as

(1′)   It is more fruitful and expedient (for certain purposes) to
          work with a language system in which numerical expressions

are class expressions of the second level than with differently
constructed language systems

(2′)   It is more fruitful and expedient (for certain purposes) to
work with a language system in which numerical expressions
are expressions of the zero level and of a special sort than
with differently constructed systems.

The claim is that whereas there is no way to settle the choice between (1)
and (2) on 'scientific grounds', the controversy between (1′) and (2′) *can*
presumably be settled as a dispute about the relative merit of two language
systems. This certainly sounds like a hopeful and optimistic advance. But
then we are told by Bar-Hillel that "there exist no generally accepted
criteria for the comparison of two language systems".[4] This makes for a
rather grim optimism! Suppose we are kind, and grant that although
there may be no purely syntactic choice criteria, or criteria for comparison,
there are indeed pragmatic criteria one could employ, in choosing be-
tween (1′) and (2′). What has one gained by the paraphrase, except some
cumbersome prose? Have the pragmatics of linguistic use been worked
out so that we have decision procedures there which go beyond those we
could have for choosing between the alternative 'ontological' beliefs
expressed in (1) and (2), in their raw, 'metaphysical' state?

    Taking it from another side: the reduction enterprise, initially intended
to eliminate metaphysical concepts (because these are not 'constituted' –
i.e. not reducible to observation statements in a systematic way – an
argument first outlined by Wittgenstein in the *Tractatus* and developed
by Carnap in *Logische Aufbau*) has had another unsuspected outcome.
The first intended victims of this critique were the so-called 'metaphysical'
terms, e.g. those held to be expendable without loss of scientific content.
In the classic case, Mach took 'atoms' to be such an expendable and
eliminable term, regarding it as reducible to sense-data terms. Russell's
argument elaborated a logical construction in which such 'inferred
entities' were to be eliminated 'wherever possible', by pristine reference
to sense-data or to sensibles (unsensed sense-data) as objects (actual or
possible) of direct acquaintance. The result is that, under the rubric of
the 'metaphysical' terms of the older positivism, there came to be included
the class of 'inferred entities', and ultimately all the so-called 'theoretical
terms' of the sciences. The salutory, though unintended effect is to implicit-

ly recognize the relation of so-called metaphysical terms to theoretical terms in the sciences. Having been made objects of the same reductionist attack, both stand or fall together, unless one introduces a clear demarcation between them. The problem is therefore to distinguish between the 'constituted', i.e. reducible concepts of science and the presumably 'unconstituted', irreducible concepts of metaphysics. Thus, two kinds of verdict hang over all higher-level terms: outright death by execution (a verdict of non-reducibility); or death by reduction (elimination by definition, or paraphrase or translation). Under the new, liberalized jurisprudence, which has grudgingly given up capital punishment, there is also a verdict short of death: the prolonged, lingering torture of partial reduction by reduction-pairs (reserved for dispositional criminals). All this is part of the political-legislative program within philosophy of science of maintaining the health of the state by vigilantly guarding the genetic purity and firmness of the reduction-base: the protocols, observation sentences, quickly-decidables, basic sentences, upon which empirical science stands. But all of this depends on identifying the proper strains which will breed true, as proper object-languages for science. How does this program now stand?

To put it plainly, the distinctions necessary for such a program have been blurred beyond recognition. With the relativization of object languages, with the separation of a so-called 'factual' from a 'formal' or 'linguistic' component of the truth of any statement, and with the conventionalism of Carnap's notion of Meaning Postulates as the systematic account of analyticity (and therefore, by some method of residues, of the synthetic), we have left empiricism in any clear sense far behind. Or at best, we have given it the status of one dogma among others. For if *anything* whatever can serve as object-language, the distinction between 'object-sentence' and 'pseudo-object-sentence' collapses, and the cherished distinction, within the domain of meaningful discourse, between the analytic and the synthetic is seriously threatened.

If the advantage of the shift from ontological *assertions* to linguistic *suggestions* is that it transforms an insoluble ontological dilemma into a manageable methodological problem, as a question of choice among alternative linguistic frameworks, one has at least to have something clearly formulated as a methodological ideal. Quine, who raised these questions sharply as early as 1950, has more recently commented on Carnap's

postulational approach to analyticity in a telling way, thus:

Legislative acts occur again and again; on the other hand, a dichotomy of the resulting truths themselves into analytic and synthetic, truths by meaning postulate, and truths by force of nature, have been given no tolerably clear meaning even as a methodological ideal.[5]

If the choice of a 'proper' object-language is a practical matter for scientists to decide on the basis of expediency, then 'empiricism' is defined in terms of whatever happens to be current in the way of object-languages; and it is a matter of going out to do the field-research in descriptive linguistics and in the sociology of science, to discover what *in fact* scientists do and say when they are at work. But this has hardly been the recommendation of even the most 'pragmatic' approaches, which remain caught within mutually exclusive demands: on the one hand, to remain true to some favored empiricist criterion of significance, in the normative program of logical reconstruction of science; on the other, to so relativize this very criterion that it vanishes into sheer description of current expediencies. The trouble here is that if one were to take the latter 'practical decision' approach seriously, then our 'pure' (reconstructed) object-languages would very soon become contaminated with the irremediable and blatant metaphysicizing that scientists engage in. The proposal for a purified empirical language of a particular sort then stands revealed as a will-o'-the-wisp, which has not been taken seriously by the larger community of working scientists, or which has been adopted simply as one dogma among others.

We may ask one more question: what if the reduction program had succeeded in eliminating metaphysical terms or concepts as empirically meaningless, and theoretical terms as redundant? Suppose that what had been preserved was something we could rigorously define as 'cognitive content'. Would anything relevant, or even necessary to science have been lost in the process? This bears on matters of practice, surely. In this respect, I think it has been shown that even where it is theoretically feasible to eliminate all theoretical terms by reduction to observation terms, it is practically unfeasible and inexpedient. Craig has shown[6] that a system containing only observational terms may be constructed to replace one containing both theoretical and observational terms. But this would require an infinite set of postulates (whether or not the postulate set of the

original theory was infinite). As Hempel and Scheffler suggest, this might answer for the epistemologist, but at a price too high for the scientist, who would (to quote Hempel), "be giving up a system of theoretical concepts and hypotheses which are heuristically fruitful and suggestive in return for a practically unmanageable system based upon an infinite though effectively specified set of postulates in observational terms".[7]

The theoretically feasible reduction thus eliminates something Hempel calls heuristic fruitfulness or suggestiveness. Now suppose we grant that theoretical terms or concepts do have such heuristic value, even if only in the weakly systematic sense which Mach would have recognized, as providing, 'mental economy', or which Duhem would have approved as providing, in his terms, 'functional economy', as a utility in science. (I will argue later that it comes to much more than this, but we may work with a minimal definition here.) The argument is that theoretical terms should be retained on the grounds of their heuristic value. But if metaphysical terms have this same value, presumably they should be retained too, despite their eliminability on other grounds. One may object that I am conflating 'proper' theoretical terms in science, with 'improper' metaphysical terms, which lie outside science. In fact, I will argue that they *should* be conflated, and that the attempt to demarcate them fails. But how might one distinguish theoretical from metaphysical terms, in respect to this question of retention of what is heuristically valuable? I think the following list exhausts all the possible distinctions:

(1)   *Metaphysical terms have no heuristic value at all, and therefore, none in science either.*

This is a straightforward, radical eliminationist or positivist view. Since, on this view, metaphysics is meaningless and cognitively empty, any use of it is obscurantist, and needs to be therapeutically excised from every corner of science, including the dark corner of 'heuristic'.

(2)   *Metaphysical terms have heuristic value, but this lies outside the domain of science proper.*

On such a view, metaphysical terms and metaphysical models may serve to interpret science for common sense, or may serve as an aid to the imagination. But these functions are extraneous to scientific work and thought proper, and serve 'only' pedagogical or psychological ends. (This

is in essence Duhem's view on the nature of mechanical models, and on the scientist's extra-scientific disposition to believe in 'natural classification'.)

> (3)   *Metaphysical terms have heuristic value within the domain of science but are to be distinguished from proper scientific terms on grounds other than those of heuristic value: namely, that they have the 'surplus content' of making ontological claims. A scientific term becomes a metaphysical one, once this ontological claim is added.*

Such a view would take any guide or program for scientific research, or anything which serves as a psychological or philosophical source of hypotheses or of models or of conjectures, or any criterion for the significance of one or another theoretical formulation, as itself a part of science (or, in a weaker sense, as relevant to science); and thus would include it in the larger economy of scientific thought, by contrast to the preceding views. There are degrees of emphasis in such a view, from that which speaks of the weak relevance, or genetic-historical relation of metaphysics to science (e.g. as preceding science) to that which speaks of the logical continuity of metaphysics and scientific theory construction, to that which sees metaphysics as one component in an inclusive *Wissenschaft* which also contains more strictly 'scientific' components distinguishable from metaphysical ones.

The last and most radical view would obliterate even such qualified distinctions:

> (4)   *Metaphysical terms are not distinguishable in any clear way from scientific terms. In effect, theoretical terms in science are metaphysical in their import, both with respect to their heuristic function (as in (3)), but also with respect to their ontological claims.*

In the foregoing analysis, the two grounds for the distinction between scientific-theoretical terms and metaphysical terms are (a) their heuristic status and (b) their ontological status, i.e. as instances of naming and referring, or as claims or conjectures that something exists. Only with respect to the second does the question of truth or falsity arise. For it is one thing to claim that metaphysics serves instrumentally (as a guide or

as a source of inspiration) in theory construction; it is quite another to claim that one or another metaphysical statement is true or false.

The parallel alternatives with respect to the cognitive status of scientific theories are the instrumentalist and realist views. The first claims only a heuristic or functional utility for scientific theories, which are therefore neither true nor false but adequate or inadequate, fruitful or not so; the second, that scientific theories *are* true or false, and that they therefore entail the assertion that something exists, or is the case, for them to be true or false about. Thus, in the latter case, the ontological or existential claim or commitment of the theory characterizes it as something more than simply an instrument of inquiry; rather, it is a claim to truth.

These alternatives are fundamental to my later discussion. Therefore, I would like to discuss some of the philosophical and historical background of their current formulation, in the context of some classical views concerning the relation of metaphysics to science.

II. SKETCH OF HISTORICAL VIEWS ON THE RELATION OF METAPHYSICS TO SCIENCE

a. *Metaphysics as Unified Science: Classical Views of a Science of First Principles*

The historical and philosophical background of this issue goes back to the joint origin of philosophy and of physical science in the cosmogonical and cosmological theories of the Ionian *Physiologoi*. Modern science developed within this classical framework, in which metaphysics was regarded as the most general and fundamental science of Being, or science of First Principles. Physical science was conceived of as that subordinate science in which one dealt with 'natural being' or with matter, and with the concepts relevant to such an inquiry. Thus it was 'natural philosophy'. Ultimate questions which were taken to be at the foundations of this natural philosophy were alternatively viewed either as fully and intimately involved in natural inquiry itself (for example, the principles revealed in natural scientific inquiry were taken to be principles of Being itself, or revelations of God's nature by means of the natural light); or they were taken as questions of an occult nature, beyond the scope of natural in-

quiry and to be referred to a 'higher study', namely, to metaphysics. (For example, when Galileo was asked about what explained *Gravity*, he referred the question to this 'higher science' and as outside the domain of natural philosophy.) Thus, in the monistically conceived hierarchy of Being, in classical and mediaeval philosophy, not only was there no demarcation between metaphysics and science, but the various sciences were ordered under metaphysics, which was 'first science' or 'fundamental science'.

b. *The Demarcation within Metaphysics*

The most radical demarcation within metaphysics itself was Descartes' two-substance theory. At one stroke, it put all non-physical questions outside the domain of natural philosophy on metaphysical grounds and introduced them instead as questions for a wholly separate science of 'spiritual substance' or mind. Never again has anyone used metaphysics so boldly to eliminate a whole set of metaphysical questions from natural science. Beyond the simple fact of this demarcation within metaphysics, Descartes' preoccupation with method, and with the epistemological grounds of his construction introduced the distinctively modern, self-conscious reflection on the nature of the world-model itself. Were such models true pictures or reflections of physical reality, or were they instead more and less adequate ways of ordering perceptual experience? Descartes' appeal lay beyond physical knowledge. The method of philosophical doubt (whose aim was really to achieve absolute certainty), rested on two grounds: (1) clear and distinct ideas, whose force of conviction in matters of belief was 'irresistable' (2) the extra-epistemological support, not simply for the adequacy, but for the truth of these ideas: namely, the argument that God would not systematically deceive us.

Thus, at the foundations of modern science and modern philosophy, the alternative emphases of realism and instrumentalism were introduced. The two views of science which develop from these alternative emphases are (1) that science is an inquiry into truth, in the sense of formulating and establishing true propositions about the world, or about nature, or natural being, (2) that science is an instrument for the anticipation of future experiences on the basis of discoverable or formulable orders revealed in past experiences (either by induction, on one view, or by invention of testable

hypotheses on another). On the first view, science is about reality, or reveals reality to our knowledge, where 'reality' is taken to be independent of our perception of it, or of our knowledge; on the second, science is about our experience, or our experimental findings, and it does nothing more than to order them effectively, i.e. in such a way that the order is taken to be a predictive instrument.

c. *'Metaphysical' Explanation vs. 'Scientific' Description: The Early Positivist Demarcation Between Metaphysics and Science*

Kant's question: 'How is Science possible?' was answered by him in terms of the necessary presuppositions required for a rational reconstruction of scientific knowledge. The incipient instrumentalism or pragmatism of the *First Critique* conceived of the usual 'metaphysical' categories as in reality nothing but necessary forms of our *knowledge* of the world, eschewing questions concerning some transcendent truth or reality-in-itself as in principle unanswerable. At best, such transcendental beliefs as we are prone to hold are regulative ideas, whose justification lies in their heuristic utility rather than in some unattainable 'truths' which they purport to assert. Combined with Kant's phenomenalism, this pragmatic emphasis was developed by neo-Kantian positivism into a distinctive doctrine in the philosophy of science.

   The methodological elaboration of these ideas in 19th-century science and philosophy of science sharply distinguished the regulative from the phenomenalist aspects of this scheme. Thus, Macquorn Rankine (in his *Outlines of the Science of Energetics*, 1855) drew a distinction between the 'hypothetical' and the 'abstractive' methods in scientific theory construction. Rankine held that the hypothetical method – what he called a method of 'conjecture' concerning the nature of a class of objects (antedating Popper in this usage) – was necessary as a first step "to put simplicity and order into the expression of phenomena before it is possible to make any progress in the construction of an abstractive theory".[8] The abstractive method then ordered only those properties of the class of objects which were 'observable', without introducing the hypothetical entities at all. Mach and Kirchoff picked up this 'abstractive' phenomenalism later in the century, and fashioned it into a striking anti-metaphysical philosophy of science. Duhem, in a brilliant but strained thesis,

then argued for the absolute separation of metaphysics from physical science, on the grounds that physics could be an autonomous science only if it abandoned all attempts to give a description of 'reality' (as contrasted with inter-subjective experience). In Duhem's terminology: if it abandoned all attempts to explain experimental laws rather than to more modestly describe them in the most economical and systematic way; i.e. by mathematical means using so-called 'concrete' (dimensional) numbers. The use of so-called 'hypothetical models' – taken by Duhem in the narrow sense of mechanical models – was decried by him as a surrender to psychological or pedagogical requirements, e.g. to the need to teach physics to engineering students in technical institutes; or more generally to the needs of common-sense imagination, rather than to the requirements of the scientific intellect. Thus, such 'hypothetical models' lay outside the domain of science proper. Always irascibly Anglophobic, Duhem held this sort of model-making against Maxwell and Lord Kelvin in particular, their *ad hoc* constructions especially distressing his Gallic *esprit simpliste*, his aesthetic-mathematical sense of system. He thus inveighed against them in the same context that he inveighed against Shakespeare, by contrast to Corneille, and Bacon, by contrast to Descartes. Kelvin's well-known remark (in the *Lectures on Molecular Dynamics*:) "It seems to me that the test of 'do we or do we not understand a particular subject in physics?' is 'Can we make a mechanical model of it?'"[9] was countered by Duhem in his critical comment on what he considered a typical 'English' example:

Here is a book intended to expound the modern theories of electricity and to expound a new theory [Oliver Lodge's *Modern Theories of Electricity*, 1890]. In it there are nothing but strings which move around pulleys, which roll around drums, which go through pearl beads, which carry weights; and tubes, which pump water while others swell and contract; toothed wheels which are geared to one another and engage hooks. We thought we were entering the tranquil and neatly ordered abode of reason, but we find ourselves in a factory.[10]

Thus, in Duhem's view, we get that association of 'metaphysics' with the notion of models, and their heuristic function, and with the notion of explanation in terms of hypothetical entities, all of which Duhem takes to stand outside of physical science proper, which should be limited to economical description of observed phenomena by means of laws.

d. *Philosophical sources of Positivism: The Evolution of Consciousness from Metaphysics to Positive Science*

Philosophical positivism had in the meantime developed its attack on metaphysics in historical-developmental terms. The irrelevance of meta-physics to science was seen in terms of a characterization of metaphysical modes of thought as the hangover of an earlier, less critical stage of human thought, dominated by mythopoeic and dogmatic religious con-ceptions. It was something mankind had gotten over, so to speak, in arriving at the state of 'positive science'. The 19th century saw a number of versions of a 'law of three stages', in which some form of 'positive' or 'rational-empirical' science represented the final stage as against earlier theological and metaphysical stages.

The most familiar form is perhaps that of St. Simon and of his disciple Auguste Comte. Although Comte did not press this as a historical or developmental thesis, but rather as a systematic one, Herbert Spencer gave it great currency in his conception of it as at least a quasi-historical thesis about the evolution of human thought.

The more systematic philosophical elaboration of this thesis derives from German speculative philosophy. One early formulation is in the Preface to the first edition of Kant's *Critique of Pure Reason*. In elabora-tions of Kant's scheme, Fichte and Hegel developed dialectical versions of this concept of stages in the growth of rational consciousness, in phenomenological and historical contexts. In Ludwig Feuerbach, this philosophical heritage finds its formulation in a form strikingly similar to, though independent of Comte's 'Law of Three Stages'. Feuerbach saw the stages (religious-theological, metaphysical, scientific) as con-tinuous in a dialectical sense: as proceeding from each other by means of criticism (on the model of Hegel's *Phenomenology of Mind*), arriving finally at a positive-scientific world-view. This view comes to full self-consciousness, when this process of development itself is rationally, i.e. scientifically understood; when it is seen that native religious conceptions are not replaced, but only transformed and given a 'metaphysical' form in theology; and when it is further realized that speculative philosophy – the tradition from Descartes, Leibniz and Spinoza through Kant, Fichte, Schelling, and Hegel – is nothing but the critically purged, most abstract form of theology itself; namely, the fantastic projection of the imagina-

tion (as in myth and religion) but here hypostatized and stripped of its sensuous 'pictorial' embodiments, and conceived in theological abstraction as a divine being, or (in the further metaphysical abstraction of this divine being) as Substance, Essence, Idea, Form, Being, etc. In Feuerbach's account, these religious, theological and metaphysical 'projections' have at their psychological basis ordinary human and natural facts – i.e. those of natural science (with regard to the physical world) or of anthropology (with regard to 'human essence' or human nature). The critical examination of the prototypical 'religious consciousness', and of its theological and metaphysical forms could be said to reveal a scientifically approachable natural and human content, by means of what Feuerbach called a 'natural' or 'anthropological reduction'. Once it becomes clear, by philosophical criticism, that "theology is nothing but esoteric anthropology", and that metaphysics is "nothing but abstract theology", then we have come to the end of traditional theology and metaphysics. In a positivist spirit Feuerbach proclaims: "My religion is: *no* religion, my philosophy is: *no* philosophy" – arguing that with the revelation of theology and metaphysics as masked and inverted expressions of natural and human truths, natural science could then replace theology and philosophy, by reinterpreting them in terms of their 'positive' and 'human' content.

On the basis of Feuerbach's critique (and in further criticism of it), Marx and Engels then characterized classical metaphysics as the passive and abstract interpretation of the world, in thought, divorced from the actual flux and press of 'living, sensuous practice' which changes the world. Thus Engels could talk, in a positivist spirit, about "the end of classical philosophy" (i.e., of speculative metaphysics), reserving only certain questions in the theory of knowledge, and formal logic as the remaining philosophical inheritance *not* replaced by science. On Marx' and Engels' view, practice guided by an ever refined and self-correcting theory was, in effect, *scientific* practice and *scientific* theory, by contrast to metaphysics in the 'dead', 'abstract', or merely 'interpretive' sense.

A surprising and insufficiently appreciated development of the Kantian-Hegelian dialectic of stages is developed by the founder of American pragmatism, C. S. Peirce, in an ingenious and novel way. In the early essay 'On the Fixation of Belief' (much read but, by and large, improperly assessed) Peirce constructs the basis of his pragmatist-empiricist theory in

terms of 'stages', which succeed each other by criticism of the 'contradic-
tions' (in terms of fatal limits) in each stage. But Peirce interprets these
stages in terms of *belief*. Here, *a priori* metaphysics is seen as the stage
which derives from the transcendence, by *rational* criticism, of uncritical,
dogmatic belief, once conflicting beliefs are permitted to confront each
other. (Popper's recent view of such criticism as the primary methodolog-
ical contribution of philosophy to science is adumbrated in Peirce's essay.)
On Peirce's account, rationality, demanding a settlement of doubt arising
out of conflicting beliefs, aims, at most, at the fullest coherence and logical
consistency in adjudicating such conflicting beliefs. But Peirce, as a good
realist, says that such coherence at its best can in no way settle the ques-
tion whether such a coherent system of adjudicated beliefs is *true*. This
may do for 'seminary minds' but not for 'laboratory minds' (to call upon
a distinction he makes elsewhere). For the answer to the question of the
truth of a system of beliefs, one has to go beyond classical metaphysics to
the scientific method of fixing belief – i.e. to that method which fixes
belief only in the practice of testing beliefs by means of the experimental
techniques of the laboratory. Thus, beliefs as 'leading ideas' or regulative
guides to action, or to scientific practice, are adjudicated by such tests,
leading to some eventual 'fated' agreement among scientists, at some limit,
infinitely distant (far off enough, at any rate, to preserve Peirce's fallibil-
ism). More recent versions of some such view of agreement 'at the limit'
(e.g. Norman Campbell's, echoing Peirce, in his definition of science as
"the study of those judgments concerning which universal agreement can
be obtained", presumably by common commitment to an ideal of experi-
mental testing of such judgments) all attest to the notion of some universal
methodological criterion for science, and for scientifically warranted be-
lief, which marks it off from metaphysics. (But, interestingly enough, does
not mark it off clearly from mathematics or logic; but this is a story to
which we shall return later.)

   In this section, I have attempted to sketch some of the historical sources
and variants of the rejection of metaphysics as either occult and lying
outside of science, or as superseded and lying behind science as vestigial
at best. In most of these views, science is extolled as at least non-meta-
physical, and at most anti-metaphysical, in one or another of the senses of
'metaphysics'. Yet, there is little doubt that in the history of science, what
one may call 'metaphysical models' (not simply 'mechanical models'

which on one interpretation are a proper subset of these) played a signifi-
cant role in scientific-theory construction, and in crucial scientific debates
over alternative theories. One has only to adduce the concepts of *matter*,
of *motion*, of *force*, of *field*, of *elementary particle*, and the conceptual
frameworks of *atomism, mechanism, continuity* and *discontinuity, evolu-
tion* and *emergence, whole and part, self-identity through change, space,
time, causality*, to suggest the budget of originally 'metaphysical' con-
cepts each of which has its clearly identifiable relevance to major con-
structs and theoretical concepts in science. One argument against such an
approach is to claim that precisely to the extent that such concepts and
frameworks *are* relevant to contemporary science, they are no longer
metaphysical; but this presumes some favored and clear sense of 'meta-
physical' to support this judgment. Surely, if one simply *defines* 'meta-
physical' as 'non-scientific' or 'empirically meaningless', or 'scientifically
fruitless', then the relation of metaphysics to science is, at best, that it
encumbers science, and therefore ought to be eliminated. But this char-
acterization of metaphysics is precisely what is at issue, and it remains to
be seen whether metaphysics may plausibly be defined otherwise.

Let me turn, then, to an examination of some recent views in phi-
losophy and history of science which consider the relevance of metaphysics
to science in what appears to be a sympathetic way, and which therefore
reject the positivist characterizations of the irrelevance or the meaningless-
ness of metaphysics. For here, it may be possible to fasten on what meta-
physics is taken to be when it is conceived as relevant to science, and
thereby, to lead to the main thesis of this paper concerning the grounds
of its relevance.

III. POPPER, AGASSI, AND KUHN ON THE RELATION OF METAPHYSICS TO SCIENCE

The views I will examine are those of Karl Popper, Joseph Agassi, and
Thomas Kuhn.

Popper's position, in short, is this: that metaphysical theories are not
literally 'meaningless', as alleged by a verificationist theory of meaning;
but rather that they are non-scientific, because irrefutable or unfalsifiable.
Thus, they *may* be true or false, but it is impossible in principle for us to
come to know this. The demarcation between science and metaphysics is

determined by the testability of the theories of science. A test therefore is an attempt to falsify a theory, and theories may be ranged in the order of their degrees of testability, with respect to how distinctively or uniquely falsifiable a theory is. The statements which are the potential falsifiers of a theory are called *basic statements (Basissätze)*, or what Popper describes as "simple descriptive statements describing easily observable states of physical bodies"[11], and elsewhere as statements which "can serve as a premise in an empirical falsification: in brief, a statement of a singular fact".[12] The irrefutability of metaphysical theories rests on the high and unrestricted universality of their existential claims. Thus, Popper writes: "A strict or pure existential statement applies to the whole universe, and it is irrefutable simply because there can be no method by which it could be refuted. For even if we were able to search our entire universe, the strict or pure existential statement would not be refuted by our failure to discover the required [instance], seeing that it might always be hiding in a place where we are not looking."[13] The refutability of scientific theories which marks them as scientific, rests on the restrictedness of their existential claims to specific finite space-time regions. The degree of testability is then a function of the degree of restrictedness. On Popper's view, a theory cannot be confirmed by positive instances; rather, it can be 'corroborated', and the degree of its corroboration is a function of the degree to which it has been tested, or the degree of the severity of the attempts to falsify it. Corroborated theories are thus those which, to one or another degree, survive such a trial.

Popper has elaborated this thesis in his major work (*The Logic of Scientific Discovery*) and in many other more recent publications. One may state the basic aspects of the thesis in terms of three constituents: (1) the notion of an empirical basis, expressed in the concept of basic statements, and in the cognate notions of acceptability, and intersubjective availability; (2) the notion of observational falsification of singular descriptive statements in the empirical basis-language. One should add here that Popper makes it clear that only statements falsify statements, and not that 'observations' do. The observational basis is introduced, therefore, as the *motive* for the acceptance of a statement, i.e. for a decision; observation alone does not therefore entail the truth or falsity of a basic statement, but only motivates its acceptance or non-acceptance. The falsifiability of basic statements then presents an additional problem

which we shall touch on presently. (3) the notion of the irrefutability of so-called unrestricted or pure existential statements (to be defined below).

The striking feature of Popper's view is his denial of *confirming* instances in science – i.e. a wholesale rejection of inductivism – and his view that scientific laws do not assert that anything *is* the case, but are rather prohibitive or proscriptive, asserting instead that something, in particular, *is not* the case. The argument here is that scientific statements require the use of 'real universal concepts', that all universals are dispositional, and that dispositional predicates are irreducible to non-dispositional ones.[14] (Popper rejects Carnap's reduction-pair solution on the grounds that it is ultimately circular, or at best leads to an infinite regress of dispositional predicates.) The argument continues, that dispositional statements transcend any finite range of confirming instances. Since confirming instances would then necessarily have to exhaust the universal concept or disposition of which they are instances, this would then amount to the rejected case of pure or unrestricted statements, which cannot in principle be exhaustively or enumeratively confirmed. Singular descriptive statements, on the other hand, are grounds for the rejection or non-acceptance of universal statements or of statements containing universals or dispositional predicates, on the logical grounds that such basic statements are the formal contradictories of universal statements, i.e. of the sort which are scientific law-like statements.

But here the argument seems to break down. Demarcation, if feasible at all, is so only if there *are* singular descriptive statements in the 'basic' language. But on Popper's own account, it turns out that such singularity is impossible to attain. Take as an example the universal statement 'All ravens are black'. Its falsifier is not the statement 'There is a non-black raven', since statements of this sort are pure existential statements; they are themselves not falsifiable because not restricted, and are therefore 'metaphysical statements' on Popper's account, and thus cannot serve as basic statements or potential falsifiers. We need therefore to restrict this statement to 'There is a non-black raven at space-time position k'. Further, to fulfil the material conditions of the basic statement, Popper specifies that the event 'non-black raven at k' must be recurrently and intersubjectively observable. Now we have presumably arrived at a singular descriptive statement, and we have a potential falsifier for '‘All ravens are black'; and given the acceptance of the statement decided on

the basis of the observation of a non-black raven at k, the universal statement is falsified.

But, on Popper's own account, the presumably singular existential statement lacks singularity unless it contains *individual concepts*, and these, according to Popper's definition, require proper names (or equivalent signs, e.g. Cartesian coordinates with a specific origin). If a statement contains only universal concepts or dispositional predicates, then it cannot be singular, cannot therefore be a basic statement, and hence cannot be a potential falsifier. Now let me list some dispositional terms, in Popper's own sense of 'dispositional': 'black', 'raven', 'space', 'time', 'location'. Even if the raven's proper name were *Sam*, and location k given in Cartesian coordinates with specific origin, this would not help. Unless the proper name can be conceived of as entirely context-free, the very propriety of the proper name lies in an infinite specification of the conditions of its applications, into which dispositions will assuredly fall like the summer rain. Popper himself stresses that the proper name will not reduce to universals, or to specification by universals, unless perhaps by an infinite specification of the conditions of application of these universals, which once again proliferates dispositions. Pure, sheer dumb ostension, at some limit of ostension, cannot save us either, for at such a hypothetical 'context-free' limit, ostension itself breaks down. (Try pointing at something to a dog, or an infant.) The replacement of names by descriptions gives still further entrance to dispositional terms, and coordinate location also requires some convention, in terms of universals, if it is to be meaningful. So Popper's singular descriptive statements, on Popper's own account of the irreducibility of dispositional predicates, are fraught with an irremediable universality, as Popper himself stresses. It remains ever a matter of the acceptability or non-acceptability of basic statements, where the very contexts of acceptance and non-acceptance are social conventions, attitudes, pragmatic and practical considerations, etc. But if the basic statements as potential falsifiers are so relativized to contexts, then have we not simply replaced the dogmatism of 'acceptance' of pure existential statements by the 'dogma of the context', or the various social dogmas of attitude and social expediency? Now this is not to argue, against Popper, that there are no instances of relatively unambiguous observation which would reasonably dispose one to agree or disagree with the statement 'There is a non-black raven at k', for so to relativize

our linguistic-perceptual behavior that *anything* can count (or not count) as an observation instance is to accede to a scholastic conventionalism which serves only obscurantism and violates scientific rationality. Arguments from illusion, and from theory-laden observation aside, we do observe ravens and writing-desks without growing old in some infinitely regressive attempt to assure ourselves. Rather, this is to argue that Popper's conditions for demarcation fail to meet *his own* criteria. Or else, what it comes to is that I may simply choose my potential falsifiers anywhere, if what ultimately counts is simply that such potential falsifiers function as motives for acceptance or non-acceptance. Even if singular descriptive statements were achievable in a disposition-free way, I would have to *share Popper's motives* for accepting *such* statements, rather than others, as potential falsifiers. And suppose I do not? What then? Is it simply that one man's potential falsifiers are another man's metaphysical statements? Is it a matter of taste? Or of social convention?

The outcome is this: the demarcation is no demarcation, since the demarcation criterion itself falls under the weight of its own dispositionality. The singular descriptive statements which Popper requires as potential falsifiers cannot be put together out of what he defines as individual concepts, without invoking what he defines as metaphysics. It is not clear that science can be demarcated from metaphysics at all, on Popper's criterion. What we have left is a continuity between higher- and lower-level terms, and a culturally and historically inherited disposition to call the higher-level ones 'metaphysical' and the lower-level ones 'scientific', and to be confused when it is not clear whether high-level theoretical terms are 'properly' scientific or metaphysical.

Popper recognizes this in his own way in considering that some unabashedly metaphysical theories, by the sedimentary accretions of scientific knowledge or by ingenious reformulation, may turn out to be components of testable theories, upon specification of their originally unrestricted existential claims (as in the case of speculative atomism, for example). Thus Popper sees the possible value of metaphysics in the tradition of critical dialectic which it engenders, from the Pre-Socratics on. Although it remains for Popper 'mythical' and non-scientific because irrefutable, still its character is anti-dogmatic, and it enters into science when "myths develop testable components". This continuity between myth, metaphysics and science, or between the non- or pre-scientific and

the scientific makes perfectly good sense. But the oddness of the falsi-
fiability criterion, and the impossibility, on Popper's account, of providing
the necessary disposition-free singular descriptive statements which he
requires for falsifiability (except at the price of attitudinal relativism) so
pragmatize the notion of demarcation that it is not clear that it is doing
any work at all. If the positivist attempt to eliminate metaphysics fails
because its verificationist demarcation between sense and nonsense fails,
so too does Popper's attempt to effect a demarcation on falsificationist
grounds. The failure is instructive; for what Popper does is, in effect, to
modify positivism, to broaden its conception of what counts as meaning-
ful (though he eschews the problem of meaning as in any way central to
his inquiry). But the old Adam of positivism is still in him, with the effect
that although he recognizes the heuristic and methodological value of the
metaphysical tradition, he cannot account in any substantive way for why
it should have this value. The simple and narrow claim for the value of the
critical method extols a thin virtue; however important it is, it remains at
most an insight into how science *may* or *ought to* work. But without a
fuller elaboration of his realism, Popper cannot give us a philosophically
adequate account of scientific truth simply on the basis of a methodology
of conjectures and refutations. Ontology may recapitulate methodology;
but it is a recapitulation worth the candle, if the philosophical theory
concerns truth and falsehood, in any sense.

Following Popper's lead, Joseph Agassi has extended the characteriza-
tion of the relevance of metaphysics to science. (Though in strict Popper-
ian terms, Agassi has claimed to refute Popper's view, for the extension
renders the demarcation criterion, as stated, false if the extension is true!)
In a recent essay ('The Nature of Scientific Problems and their Roots in
Metaphysics'[15]) Agassi holds that metaphysical theories stand or fall
(within the framework of rationality) by virtue of the refutability of the
scientific theories developed as *interpretations* of them. Thus, although he
seems to maintain the demarcation criterion in Popperian terms, he
suggests a test for the acceptance or the rejection of metaphysical hyp-
otheses which introduces a very different sense of 'demarcation', to wit:
criticizability, in a weaker sense than *refutability*. Agassi bases his view
on a thesis about the history of science. He claims that, as a matter of
historical fact, the criterion for the choice of scientific problems is their
relevance to then-current metaphysical problems, and not, as Popper

claims, their high degree of testability, or refutability. He speaks of meta-
physics as a "coordinating agent in the field of scientific research" in the
sense that, in the choice of scientific problems, the criterion of signifi-
cance, or of importance is metaphysical relevance. There is, according to
Agassi, a kind of crucial experiment for metaphysical theories therefore,
which he describes thus:

Two different metaphysical views offer two different interpretations of a body
of known fact. Each of these interpretations is developed into a scientific theory,
and one of the two scientific theories is defeated in a crucial experiment. The
metaphysics behind the defeated theory loses its interpretive power and is then
abandoned. This is how some scientific problems are relevant to metaphysics;
and as a rule it is the class of scientific problems which exhibits this relevance
which is chosen to be studied ... metaphysical ideas belong to scientific research
as crucially important regulative ideas.[16]

In this context, Agassi asserts that

scientific physics belongs to the rational debate concerning metaphysical ideas.
Some of the greatest single experiments in the history of modern physics are
experiments related to metaphysics. I suggest that this relevance to metaphysics
contributes to their uncontested high status. And yet, I contend that the meta-
physical theories related to these experiments were not part[17] of science.[18]

Thus, Agassi modifies the demarcation criterion of Popper, adding
that the goal of scientific research is not, as Popper says, the finding and
testing of highly testable hypotheses, but rather the finding and testing of
metaphysically relevant hypotheses. He thus seems to turn the demarca-
tion criterion into a heuristic criterion of rational choice among theories
to be seriously entertained. Where Popper had proposed, in a purely
methodological context, that the move which connects observation to a
good theory is not the inductive move from observation to theory, but is
rather to be made by proposing *any* theory and then testing it[19] – the
criterion of choice as to which among the testable theories to test is, of
course, the degree of falsifiability or refutability of the theory: i.e. always
choose the most severely testable theory – Agassi proposes that meta-
physics operates heuristically to control the choice among testable theories,
with respect to which of these will actually be tested; and also, in some
sense, metaphysics works to generate those theories which will be re-
garded as significant in the 'problem situation'; this, by the presumably
clear move of "developing a scientific theory" out of a metaphysical one.

In another paper, Agassi describes metaphysics as "prescribing pro-
grams for future scientific development" and says that such programs
are "neither true nor false but commendable and condemnable".[20] Yet,
according to Agassi, the metaphysics which prescribes such programs is
not only heuristic, but is either true or false: if true, then commendable,
and if false, then ordinarily, but not necessarily, condemnable. For it
turns out that a false metaphysics may also issue in a commendable
program. (Faraday is Agassi's case in point here.) But then the commend-
ability and condemnability of programs has no immediate bearing on
the truth or falsity of the metaphysics which prescribes them, plainly, but
only on its heuristic value. The question arises then: what *does*? If the
demarcation criterion is to be preserved then either it is the case that
metaphysical theories are true or false independent of their refutability
(and Agassi does introduce the notion of 'false though irrefutable') or
there must be a way to decide on the truth or falsity of such theories
entirely apart from refutation (in the sense of falsification by empirical
test). At the limit, we should have to say that metaphysical theories, since
they are meaningful, are either true or false, but that there is no way to
decide this. What Agassi proposes therefore abandons the whole demar-
cation issue. But then it remains unclear in what sense such theories are
"not part of science". Agassi proposes adopting Popper's generalized
view of rationality as including not only empirical refutation in science
but also critical refutation in philosophy. Presumably, it is where critical
refutation fails that indirect corroboration and refutation of metaphysical
theories as either true or false may take place on the basis of empirical-
scientific outcomes. Agassi clearly goes beyond Popper in this. But his
examples need examining. He writes:

Marx's prophecy [?!] about the geographical location of the socialist revolution
has been refuted by his Russian followers; and this *amounts to* [my stress, M.W.]
a refutation of his materialism since it *entails* [my stress, M.W.] the valuelessness
of imaginative ideas.[21]

Ignoring the rhetorical overstatement (or just plain perversity) here,
we may reconstruct Agassi's scheme in the following way

$$((M \to P) \cdot \bar{P}) \to \bar{M}.$$

Perhaps this is too strong, if we interpret '$\to$' as entailment. Agassi would

claim that the metaphysical theory (Marxism) is *interpreted* so that P (say, the proposition concerning the location of the revolution) 'follows' from it, in some informal sense of 'follows'. Still, Agassi would claim, the metaphysical theory is directly refuted *empirically*, without the need to go to a scientific theory from M. This goes beyond Popper, clearly, if the Popperian demarcation claim is that the metaphysical theory, is, *ex hypothesi*, irrefutable; and becomes refutable only on its interpretability as a scientific (i.e. falsifiable, or refutable) theory. The Popperian scheme would look like this:

$$(((M \to T) \cdot (T \to P)) \cdot \bar{P}) \to \bar{M}.$$

What demarcates metaphysics from science here is the first conjunction sign, which does no more than to claim the interpretability of metaphysics by some appropriately falsifiable theory. But this is no more than to give metaphysical terms the status of theoretical terms of high order *within* science. And then it makes little sense to call metaphysics non-scientific, as Popper presumably would. But it would make no more sense to call metaphysics non-scientific as Agassi does.[22] Also, one may call into question the notion of 'crucial experiment' to refute a metaphysical theory (e.g. 'materialism' here) when it is proposed in such a simplistic and erroneous way. Agassi finally has to show what the relation is between the commendability of a program, and truth, before it can be clear what it means to say that a metaphysics can be false.

Thomas Kuhn, in *The Structure of Scientific Revolutions*, offers another thesis about the relation between metaphysics and science. He proposes that in any mature science, there have been developed *paradigms* which he defines as "universally recognized scientific achievements that for a time provide model problems and solutions for a community of practitioners". Kuhn regards these 'shared paradigms' as the fundamental 'units' for the study of the history of science. They constitute more than mere modes of practice, but are rather complexes of modes of thought and operation shaped by a "nest of metaphysical and methodological commitments".[23] He writes:

Effective research scarcely begins before a scientific community thinks it has acquired firm answers to questions like the following: What are the fundamental entities of which the universe is composed? How do these interact with each other and with the senses? What questions may legitimately be asked about such entities and what techniques employed in seeking solutions?[24]

The paradigm serves therefore to define cognitive significance (the 'legitimacy of questions') and ontological commitment (what entities exist) and *thereby* serves as a heuristic for the formulation of problems and the choice of methods for their solution. The paradigm, further, drastically restricts the vision of the scientist, so that this focus permits intensive investigation of a "small range of relatively esoteric problems", which would otherwise be unimaginable.[25] A 'scientific revolution' occurs only when the paradigm has resulted in so fine a mapping of its domain that anomalies, when they occur, become sharply realized. Such anomalies create crises in science, which give rise to revolutions, as 'transformations of vision', by the construction of new paradigms. "What were ducks in the scientist's world before the revolution are rabbits afterwards", writes Kuhn.[26]

Such a revolution involves, therefore, a shift in metaphysical commitments, as a feature of the paradigm-shift. The elements of the paradigm are so intimately interrelated that according to Kuhn it is not clear that one can make any categorical separation between scientific fact and scientific theory. And since both are paradigm-bound, Kuhn questions whether Agassi's version of the separation of science from metaphysics is finally realizable. In a comment on Agassi's paper, Kuhn asks: "... can we still use the distinction between metaphysics and science truthfully in our analysis?"[27]

In summary: Kuhn's view is that metaphysics, in the sense of some world-picture into which the domain of a science fits, is indistinguishably bound up with the formulation of the very criteria of scientific significance in any mature science; therefore, with the very notion of fact, and with the conditions of theoretical significance. No demarcation is possible, therefore.

Thus, we have three views – Popper's, Agassi's, and Kuhn's – all of which recognize the relevance of metaphysics to science: Popper, in terms of the original myths which, though irrefutable and non-scientific, may develop testable components, and which serve to establish the tradition of criticism and anti-dogmatism in which the development of science becomes possible; Agassi, in revising Popper's view and claiming that there is a sort of weak but direct refutability, on empirical grounds, of metaphysical theories; but retaining all the ambivalence of (a) holding metaphysics to be primarily regulative in proposing programs to science,

which programs as such are neither true nor false, but commendable or not, and (b) also holding metaphysics to be true and false on some *other* (unspecified) grounds; and finally, Kuhn, who finds any scientific matter of fact so embedded in a theory-and-metaphysics-laden matrix, that he disclaims any categorical demarcation between these elements.

Certainly, this is a far cry from the positivist anti-metaphysical stances of the past. But are these the philosophically cogent, rational and broad-minded alternatives to a narrower empiricism that they appear to be? What makes the issue unclear (and what is after all central to its resolution) is that 'metaphysics', as we see from these alternative formulations, has the unsettling and Protean capacity to change and adapt itself to fit and support each view comfortably. If we look to see what Carnap takes metaphysics to be, and what Carnap takes meaning to be, then (insofar as either of these is clear) it follows that metaphysics is 'meaningless'. Popper's 'metaphysics' is irrefutable *ex hypothesi*; and just as Agassi's 'metaphysics' is constituted to regulate and propose programs, so too, Kuhn's 'metaphysics' is the sort that will fuse with the organically inter-penetrating elements of the paradigm. Either 'metaphysics' is so extra-ordinarily rich a concept that it can serve all these uses equally well, or the flexibility of the term is so great that it can be defined so as to suit it to serve one or another thesis. One man's metaphysics may be another man's science (or his nonsense) but only if metaphysics is a sort of philosophical wanton, an imaginative *fille de joie* able to bed down with, and satisfy the most diverse philosophical libertines.

In all three views, metaphysics is taken as heuristic to one or another degree – from the point of view of its self-critical tradition, or from its suggestion of models as anticipatory constructs which may generate scientific theories, or as directly refutable metaphysical hypotheses, or as a set of operative presuppositions concerning cognitive significance and ontology. But all this is plainly acceptable, because harmless, or inno-cuous, or vacuous. On the one hand, the history of philosophy offers no evidence that metaphysics has a viable 'essential' definition, so one cannot insist that the true faith has been violated. On the other, all the players stay within the rules of their own games reasonably well. As each one defines metaphysics, it does what it is cut out to do, context by context, as one might expect. If one were content with small favors, one could argue that, after all, the tradition of systematic philosophy is saved, if

not in its pristine form, at least in some qualified and assimilated form.

My objection to this small-scale payoff is that none of these views ask the deeper question: How is it that metaphysics does operate heuristically in science? Agassi explicitly says he does not know *why* significant events in the history of science should be metaphysically relevant, but only that he has found it almost always to be the case.[28] Kuhn's thesis carries little normative force except perhaps as a proposal to study the history of science with due respect for philosophical and sociological aspects. Metaphysics serves Kuhn, therefore, as an object of description only, to be taken into account in characterizing the 'unit' of historical study. Popper recognizes the problem as Kant's question: "How is science possible?", but his demarcation prevents him from getting at the roots of it.

The problem we inherit from the epochal historical-analytic studies of a Whewell, a Duhem, a Meyerson, a Cassirer, is: "Why should metaphysics function as it does in science? What explains its heuristic value?" The question, in short, is: "What is metaphysics?" For only with this question can we approach the issue of why metaphysics has heuristic value, and go beyond a descriptivist interpretation which, however useful, only tells us what we should already have known, if the positivist scales had not adhered to our eyes for so long; namely, that in point of historical fact, metaphysics *has* served this function. This is not a new discovery, and it is incredible that the statement of this almost trivial fact should occasion as much response as it does these days, especially, when it is stated as no more than a fact. The task at hand is not to describe the relation between metaphysics and science, but to explain it. Popper fails to answer this question; Agassi and Kuhn fail to ask it. In Kuhn's case, with respect to his debt to Polanyi, we might guess that the prospective answer lies in Polanyi's account of what he has called 'heuristic vision' (which Polanyi himself suggests is like what Kuhn means by 'paradigm').[29] If it does, then I would judge that *this* alternative to positivism lies in the direction of a total abandonment of the rational objectivity of science, and substitutes instead a voluntarist obscurantism. In the cases of Popper and Agassi, the realist *intentions* are clear (e.g. in Popper's 'Three Views of Human Knowledge'), but their theoretical and philosophical support is not. If this were pursued, perhaps the relevance of metaphysics to science would not simply be noted and characterized, but explained as well.

Unless a more adequate account is forthcoming, of something we

might call 'scientific realism' which does not simply take metaphysics into account, but copes with it seriously, we are left with the alternatives of Polanyi-Kuhn fideism; or else, with a reform within the framework of positivism – a piecemeal reappropriation of metaphysics, by an emasculatory process of redefinition and re-interpretation to suit one or another weakened criterion of cognitive significance – effectively, a process of letting metaphysics in the back door, through the 'service entrance', so to speak.

IV. PROVISIONAL THESES ON METAPHYSICS AS HEURISTIC

In this last section of the paper, it is therefore not my intention to show *that* metaphysics has been heuristic for science; for that it was so, historically, is too plain to need demonstration, and there is a rich literature to establish this claim: in Mach's own historical-analytic works on the science of mechanics and on the conservation of energy, in Duhem's history of cosmological doctrines as in his other works, in Anneliese Maier's work on the precursors of Galileo, in the works of Meyerson, Koyré, Metzger, Lasswitz, in Burtt's and Lovejoy's treatises on the history of scientific ideas, in Jammer's recent work on the concepts of space, force and mass, Hesse's on forces and fields, Čapek's on the conflict of mechanist and process views in physics, and in many others.

Nor is it my intention to assess fully the extent to which this function of metaphysics has been of positive value to science, or has been pernicious: this is a question of detailed historical study and cannot be given a simple programmatic answer. Nevertheless, it is clear that in a fundamental sense, I take metaphysics to be a condition of any scientific understanding whatever, and in this sense clearly take it as of positive value to science. But even so, it should not be problematic that it can be pernicious as well: by loose analogy, breathing may be considered a necessary condition of life, and of positive value; but breathing noxious gases may just as well kill one.

Rather, I want to ask: "Why does metaphysics function heuristically in science? What explains it?" For in showing why it has served this function, I think we may show why it should be of current interest as well. In order to inquire into this, let me first set out some definitions and considerations.

First: the term 'heuristic' has the general connotation of a guide, an aid to some end, a source of suggestions as to how to proceed – either in the formulation or in the solution of some problem; or where formulation and solution are closely related, as they often are, in the setting of a frame-work for this joint operation. If a heuristic for science, then the end of such a heuristic (in science) needs to be made clear; for on one or another view of what this end is, some concept of what science is is already implied. If, on the one hand, we take science to be an instrument for successful prediction, then our heuristic will have to function to that end. It will be a heuristic for successful practice. If, on the other hand, we can draw the distinction between the *practice* of science and scientific *under-standing* (as I think we can) then we require a heuristic of a different sort. I would like to distinguish therefore between a *heuristic for practice* – for example, an algorithm, or a paradigm of experimental procedure – and a *heuristic for understanding* – which I take to be a conceptual or theoretical model. The distinction may be further sharpened in this way: scientific practice is clearly concept-ordered, and entails the *use* of concepts which the scientist learns in his apprenticeship as the tools of his trade. Such practice I would call *intelligent* but not yet *rational*. Rational practice entails something more: not simply the use of concepts, but the self-conscious or reflective use of concepts; i.e. the critical attitude towards scientific practice and thought, which constitutes not simply scientific knowledge alone (which is its necessary condition), but the *self-knowledge* of science, the critical examination of its own conceptual foundations (or minimally, the readiness to bring such examination to bear at crucial times). This is really Kant's distinction, and defines what he set out to do, as he describes it in the preface to the first edition of the *First Critique*. If we can show that metaphysics is effectively the way in which this critique is systematically carried out, then the examination of the grounds of why metaphysics functions as heuristic for science is at the same time an examination of the grounds of rational scientific practice.

But in what way does metaphysics serve the function of aiding the understanding, the self-knowledge of science? Let us grant that under-standing is constituted by an examination (a critique) of the conceptual foundations of science. One might very well object that this is the general task of philosophy, or specifically the task of philosophy of science, and not of metaphysics. If this much is granted, then my thesis is that this, in

fact, *is* metaphysics. That is to say, any systematic critique of the conceptual foundations of science, if it is sustained, and not simply *ad hoc* explication of this or that concept, ends up coping with those questions we would recognize as metaphysical. If one does fundamental criticism of the set of concepts, of working theoretical assumptions, or of the models of a science, then there arise such questions as those concerning the characterization of causality, of the entities which the science countenances, of the fundamental relations in terms of which the scientific system operates, of the atomicity or continuity of the phenomena, or of the 'space' of that system. Whether we take this as a physical geometrical 'space', or a sociological or psychological 'space', the point is that such 'spatial' conceptions, in different sciences, have interesting common structural properties. In short, systematic critique, if it is sustained, eventuates in an abstract model of the science itself, which functions just as models in the sciences do: namely, to aid us in understanding the science.

But here, metaphysics (or the knowledge of metaphysics, i.e. of the classical metaphysical systems in the history of philosophy) seems to me clearly relevant, and clearly has a heuristic function. I would argue, that with respect to the most general features of such conceptual models of the sciences themselves, metaphysics represents the basic theory of such models, and that classical metaphysical systems (theories), or their variants, may be mapped onto such models of the sciences. That is to say, the representation of the structure of a science is a model (an interpretation, a mapping) of a more general and abstract theory of structure, which I take a metaphysical system to be. Thus, the study of metaphysical systems provides the most rigorous, systematic and abstract exercise in such construction.

Further, if we study metaphysics in this way – as an enterprise of 'conceptual exploration' as Gerd Buchdahl has suggested – then we are in a position to separate the rational elements of classic metaphysical systems from their dated, literal content which has often been repudiated by subsequent scientific knowledge. Then the history of alternative metaphysical systems reveals itself as a rich heritage of *theories of structure*, in which the essential features of theoretical construction are set forth in the most general way. Further, the history of metaphysics reveals to us models of criticism in terms of which conceptual systems of great generality were constructed, torn down, modified, replaced, in which primitive concep-

tual formulations were developed by criticism, by elaboration, into the
most highly complex and systemic theoretical constructs. (Merely to
mention the Pre-Socratics, and the great alternative schemes of Demo-
critus, Plato and Aristotle, suffices to make the point.)

Metaphysics, as the most general and abstract such theoretical enter-
prise, reveals the structure of theories most clearly. The main features
which are thus revealed, in any metaphysical system worthy of study, are
(1) *systematic structure*, (2) *reference*, (3) *abstraction*. In this sense (which I
propose to examine) metaphysics provides the most general schema for
conceptual models. And since I am going to argue that we come to *under-
stand* only in terms of such conceptual models, metaphysics becomes a
heuristic for understanding, by providing the most general and abstract
account of the conditions under which anything whatever comes to be.
understood. Now such a broad claim might suggest that almost anything
whatever is 'metaphysics' insofar as it functions in the context of under-
standing; and one may easily trivialize this thesis by asking whether it
requires a metaphysical system to come to understand, for example, the
meaning of the word 'cat'. But it is something almost as perverse as this
that I am recommending. My argument might run that if we pursue in a
fully critical, systematic way, with due attention to problems of reference
and abstraction, such innocuous examples as the one given, then what we
will end up requiring for this understanding will be a metaphysics. As
any philosophical pedagogue knows, the starting-point for metaphysics
may be almost anywhere; it is the conditions of proceeding from this
starting-point that count.

I would argue, parenthetically, that the dispensability of metaphysics
is equivalent to the dispensability of rationality: that metaphysics is the
practice of rationality in its most theoretical form; and that therefore,
what metaphysics underwrites is the *theoretical* understanding. But I
would argue further that 'theoretical understanding' is as redundant a
phrase as 'round circle'. This may be taken as no more than a suggestion
as to how to use the term 'understanding', but if it is this, then such sugges-
tions are not simply semantic fiats or arbitrary conventions, but entail philo-
sophical arguments. Peirce's argument on the dispensability of metaphysics
is apropos here. In his 'Notes on Scientific Philosophy', he wrote:

Find a scientific man who proposes to get along without any metaphysics – not

by any means every man who holds the ordinary reasonings of metaphysicians in scorn – and you have found one whose doctrines are thoroughly vitiated by the crude and uncriticized metaphysics with which they are packed. We must philosophize, said the great naturalist Aristotle, if only to avoid philosophizing.[30]

What is wrong with avoiding or evading or dispensing with the study of metaphysics, therefore, is that it leaves one at the prey of whatever one has uncritically picked up along the way, and to which one hews by way of unconscious or conscious dogma.

Now, having made this broad a claim, I need to add that not everything is a metaphysical theory, and that among metaphysical theories there are good ones and bad ones. Popper's 'pure existential statements' are hardly 'metaphysical', and Agassi's example of Faraday's 'metaphysics' – namely 'All is force' – is no more than a crude metaphysical hunch, but hardly a metaphysical system, as it stands. Bad metaphysics, if one may briefly characterize it, is sloppy metaphysics, lacking rigor in construction, lacking richness in characterization of its entities, or lacking originality, merely producing bad copies of good originals. A bad metaphysics is therefore one which fails structurally, fails referentially, and fails abstractively. If we adopt the criteria I proposed (namely *systematic structure, reference* and *abstraction*) then a lack of structural rigor or of clear reference (a muddling or inconsistent or opaque characterization of entities, for example) marks a weak metaphysics, as a lack of generality marks a narrow one – for example, a 'metaphysics' of humor, or of food, or of heroism fails by the paucity of the domain over which abstraction operates.

Now many scientists are full of metaphysical hunches, but not many scientists, in my experience, can follow a metaphysical hunch across the street. The ones who can, I suspect, are (or can become) outstanding theoreticians. I would go so far as to say that such scientific theoreticians are, in a distinctive and critical way, elaborating and continuing the leading ideas of one or another metaphysics; and having abandoned the problem of demarcation, I find no difficulty in this. A rational theoretical science is continuous with the tradition of metaphysical theory-construction, falling short only by the departmentalization of its scope. But even here, such a science reveals structure and systematizes knowledge in such a way as to permit the enlargement of its original scope. Contemporary biology is no longer indifferent to physics, and the vast unification of

chemistry which proceeded from the atomic models of quantum physics is well known. While the problems of general scope only *begin* with such achievements, their direction is clear. It is precisely that marked out by classical metaphysics: a rational – that is a unified, coherent and critically appraisable – world-picture; in short, a model of reality.

Here, perhaps, the centrality of reference becomes clear. It would seem that in talking of metaphysics as a model of *understanding* we preclude conflating it with a model of *reality*, for these are certainly not the same. And yet, what is involved in attempting to understand anything whatever is to take something as real, and propositions about it as true. In a trivial sense, but one which does not simply confuse what we understand and know with what there is, any significant understanding, any scientific claim to knowledge, no less than any ordinary epistemic claim, proposes that something is the case, that something exists. Indeed, theory is a rational technique for making and testing proposals of this sort.

Even if what is taken to exist, or what is taken as the reference of the claim is an impoverished or severely restricted domain – e.g. the numerical data of an experiment, or the 'observed phenomena' – this too involves ontological commitment. When Russell called for the reduction of physics to sense-data, these new, queer entities were taken by him as 'physical existents'. And when Quine economized ontological talk with the definition: "to be, is to be the value of a variable", or when Carnap defined ontological commitment as an 'internal question', simply relative to choice of linguistic frameworks, the inevitable claims of reference made themselves felt. Even in the case when we 'understand' what is false, the significance of this very assertion drags ontology along with it (and ontological problems in its wake – e.g. as to whether false but meaningful assertions 'refer', say, to the null class as their extension, or are simply such as fail of reference, etc.).

Here too, metaphysics imposes this condition of reference on systematic structure, and serves as the arena in which such problems are grappled with. In a sense, then, metaphysics imposes the responsibility of dealing with questions of truth and falsity in a rigorous, systematic way. The various pragmatic-conventionalist and realist arguments in philosophy of science, concerning the cognitive status of theories, do not differ on the centrality of the question of truth and falsity, but serve rather to articulate alternatives. One may distinguish, therefore, between

questions concerning the *definition* of truth, or the nature of reference, and questions concerning *criteria* of truth, in terms of which we judge that the truth-conditions of one or another definition have been fulfilled. And here, of course, there are clearly alternative arguments.

But such arguments make sense only as themselves critical elaborations of one or another metaphycisal scheme, some responsibly systematic account of the relation of our practice to truth, of our knowledge-claims to reality. Calling this account something other than, or less than meta-physics simply mistakes what metaphysics is all about, and leads to that peculiar ahistorical and anti-historical attitude towards philosophy which has dominated philosophy recently, and has seriously impoverished it.

Thus, I am talking about metaphysics as that heuristic which serves the end of helping, guiding, suggesting how the scientist comes to *under-stand* what he is doing, and not simply how he comes to *do* what he is doing. Thus, it is the heuristic which guides theory-formation insofar as understanding in science is a matter of theoretical explanation and in-volves the use of theoretical models.[31] This takes the aim of science to be *not* successful prediction *tout court*, but that rational understanding which makes the rational practice of prediction possible (as against random or 'lucky' guessing).

Still, why *should* metaphysics serve as heuristic? What accounts for it? I would like to suggest here that it is because it has its roots in the deepest and most pervasive primary experience of understanding that we have, in that most common means of coming to comprehend, and of *learning* to comprehend: the *story*. Let me digress to make this clear.

Etymological arguments are always a shaky business, at best, in phi-losophy, and should be appealed to for no more than their suggestions of interesting ways to consider questions. In this case, it is worth the risk. The Greek term *logos* in its philosophical usage has its roots in an older verb *legein*, which means 'to read' or 'tell'. This in turn goes back to a root notion about the *common* or that which is shared in, or participated in by all. This serves only to introduce a common point: that the earliest sharing we do, the first commonalty that we establish, in any rational capacity, is in discourse. Speech humanizes us, makes us conscious members of the same species. We come to know by coming to speak and to listen and to understand each other's discourse. Whatever innate con-ceptual capacities there may be, their payoff is in communication by

means of language. And the first *formal* structure, beyond expressive or gestural discourse, by means of which we are introduced into the human universe of discourse as full-fledged members of the species *homo sapiens* is the story.The point is trivial but not generally remarked, nor sufficiently appreciated outside the contexts of developmental studies in learning theory, or studies in the origins of literature and drama.

The story recounts a non-present series of events. It represents the actors by naming and characterizing them (as 'Big Bad Wolf' or 'Chicken Little'). As children, we are delighted by the recognition of familiarities, and equally delighted with the novelty of the unexpected yet comprehensible event, in that as yet undifferentiated and vivid limbo between imagination and experience. Language, the chief artifact of man (if not, as recent arguments (e.g. Lenneberg's) in linguistics hold, a genetic species-specific 'capacity') becomes the means for the representation of a *structure* of events and of entities, whose very names and characters tell us what to expect of them. It becomes, in this function, the school of our understanding: *we come to understand when we understand the story*. And every generation, and every culture recapitulates this initiation. It is at the basis of our primary experience as human beings in a community of others we take to be like ourselves; and identification with the characters, the actors, the events, the affective states represented in the story becomes the earliest articulation of our awareness of ourselves as members of a species.

But what is the story (or its elaboration, the myth)? It is plainly a conceptual structure or a model, whose main features are (1) that it is more or less well connected and plausible in its parts, i.e. has some *systematic structure*, (2) that it represents certain personages or actors taken to exist and certain events taken to occur, during the suspension of disbelief or even the literal belief which accompanies storytelling, i.e. it exhibits the features of *reference*, and (3) of necessity, selects only those aspects of the character of the heroes or villains, and only those events of the daily routine which are significant to the story, i.e. it is an *abstractive* account. The story thus represents the first and most pervasive conceptual model in our experience. Our lives are thereafter filled with the telling of tales, each of them calculated to make the most out of the most important features. Our early education gives us our mathematics and our science in story form, in the concrete imagery of 'John went to the store and

bought three apples...', as our later education gives us the abstract imagery of the logical, mathematical or scientific 'story' with *its* entities and their 'characters', their relations and the rules of their concatenation, and the outcomes to which these lead.

Thus too, our behavior and beliefs are shaped by story-paradigms, by parables, histories and myths, as these enter and become thoroughly enmeshed in our very perceptual experience itself, and in our judgments of 'direct' experience. The 'stories' articulate us, as we articulate them. Our common-sense concepts embody these models in a general form: they are the implicit rules of 'normal' behavior, of 'normal' expectation and belief which a combination of inheritance and acquisition, of observation and imitation, and a long exposure to story-paradigms have shaped. We do not shed our first skins easily. The use of the conceptual model as the instrument of understanding is deep-grained, common, ubiquitous and, I believe, ineradicable. What we can do, however, is to become reflectively self-critical, in taking these models, and common-sense itself therefore, as objects of criticism. But to make them criticizable in the first place requires articulate criteria of their adequacy; and in the second place requires that structural, referential and abstractive features of such models be made explicit to the degree that they can be subjected to criticism, in its manifold dialectical and empirical aspects. This seems a bootstrap operation, for if, as I said, the use of models is indeed ineradicable, then the very procedure of criticism *itself* requires a meta-model. And it is the critical procedures of metaphysics, and of the history of metaphysics which provide the meta-models here (i.e. the models of story-telling, or of model-making).

Talk of models often suffers from a simple confusion, easily set right. As I have been using it thus far, the term 'model' connotes an abstract and conceptual representation of some concrete domain taken to be the 'object' or the reference of the model. The manner of speaking which logicians and mathematicians have adopted is the inverse of this, so that 'model' in this alternative usage connotes an interpretation or a concretization of some abstract or uninterpreted formal system, or calculus. Just as the former usage has its virtues in suggestions of the origin and genesis of models as abstractions from experience and practice, so this latter usage has its virtues in suggesting the free character of logical or formal construction – akin to the free creation of stories – and the sub-

sequent interpretations or embodiments which such constructions may take.

There are problems of ontology even for so-called 'uninterpreted' calculi, which I will evade here. But a point about the relation of metaphysics to logic may be made, without a fuller elaboration of this argument. The conceptual models of science vary in character: some are of extremely limited scope, and are *ad hoc*; others are of great systematic scope, encompassing a whole field of inquiry, or systematizing several fields of inquiry at once. All of them exhibit the features of systematicity or structure, reference and abstraction. All of them are, in the logician's manner of speaking, *interpretations* (or *mappings*, or *models*) of one or another more abstract or formal structure, in the sense of being applications of such a structure to a specified domain. The interpretation, in effect, puts values for variables in the formal system. Now if we take as a *relatively* unspecified formal system one or another metaphysical scheme, we may find various interpretations of it. Thus, for example, *atomism* constitutes such a relatively abstract metaphysical structure, which has found its interpretations variously, in physics (perhaps the original source of its generalization as a scheme), biology, psychology, sociology, linguistics, and even (with the hedonic calculus) in ethics. It has its counterparts as well in common-sense contexts, in fairy-tales and myths, and in theology. In this sense, one can relate the whole discussion of uninterpreted calculi and their interpretation to the discussion of the 'interpretation' of metaphysical systems, though these latter are not calculi in any formal sense.

That metaphysical systems are thus formalizable, however – perhaps, in a weak sense, even axiomatizable – suggests some significant relationship to logic. We need this *caveat* at least: 'uninterpreted', for a metaphysical system, means uninterpreted for a specified domain – *relatively* uninterpreted therefore. For every metaphysics presumes to *be* an interpretation, in itself, of world-structure, or more generally, of the structure of what is. No metaphysician who takes himself seriously proposes merely an alternative purely formal scheme devoid of ontological intentions. Every metaphysical system of any significance lays claim to being rational (well-formed, consistent, articulate, criticizable). But beyond this, most classical metaphysical systems tacitly claim to be true (except in the rare instances in which they are explicitly set out as imaginative hypotheses,

or as constructions to which alternatives are also offered). But even in the conjectural case, a metaphysics is at least put forth to be entertained as a model of world-structure, however tentatively. Thus, a metaphysical system may be seriously offered as a true and adequate description of reality. But at the same time, it serves as a model of the form which such true and adequate descriptions should have. And in this sense it is the minimal model of every scientific theory, both as to form and intention. As a 'first-level' model of reality, it purports to be true or adequate. As a 'second-level' meta-model, of the form of systematic representation, it serves its heuristic function in theory-formation, and also serves to suggest various interpretations, as 'first-level' models, or alternative mappings of it, in one or another scientific domain.

With respect to systematicity and to questions of ontology as well, we may suggest that logic is not simply akin to metaphysics, but is a part of it. It is certainly not the whole of it, then, but perhaps an attempt to formulate the minimum common part of any metaphysical *schemata*, which then may differ among themselves as interpretations of a logical scheme. This makes all metaphysics subject to logical criteria; and conversely, it lays a metaphysical burden upon the philosophy of logic, upon the discussion of the status of logical criteria themselves. This is no longer 'logic without ontology', but rather relates abstract theories of structure and reference to metaphysics. Metaphysical interpretations of logic, e.g. platonist or nominalist interpretations, stand to logic in a relation similar to that in which scientific theories stand to metaphysics. Thus, while a 'pure' logical structure may have only the most tenuous connections with empirical and testable components of a scientific theory, the 'logic' of a scientific theory is no less a constituent of the theory than the metaphysics is.

The fly in the metaphysical ointment is that alternative metaphysical theories cannot *both* be true and adequate, when incompatible. They may both be considered true, if complementary, but then neither is adequate alone, and they are only mutually adequate. This would yield a seamed rather than an unseamed world-picture, or a frank dualism. Or it may be that one metaphysical theory is true and adequate, and the other false; or that both are false. (Mary Hesse has recently characterized the logical alternatives in the situation concerning wave and particle models in quantum physics in a similar way: namely, (1) neither is true (or necessary) as

a distinct model, (2) both are true (and necessary), (3) only one or the other is.[32]) The remaining alternative (which lies outside the *tertium non datur* of being either true and adequate, or false and inadequate) is to view metaphysical theories as being neither true nor false, adequate or inadequate. Hesse's (1) may be interpreted this way (*no* models, neither these two of wave and particle, nor any others, are either true or necessary); or it may be interpreted as demanding a *third* alternative, on the view that a different model entirely is required for quantum theory, to which 'wave' and 'particle' are at best partial approximations, or are at worst totally irrelevant and misleading or false.

Since my argument is that scientific understanding requires a model, then the 'no models' alternative is hereby rejected. We are left then with complementary models (or metaphysical theories), or incompatible ones, in which either one is true and the other false, or neither is true.

But why should incompatibility bother us, or even complementarity (as it assuredly does some of us, Bohr and Rosenfeld excepted)? Why not take metaphysics serenely as an exercise in concept formation, adopting any rules we wish? Is it because, as Mary Hesse says in a recent discussion, incompleteness "offends against a deep-seated conviction that reality is ultimately one, and that there must be a unity in the fundamental concepts of our theories"?[33] If there is such a deep-seated 'conviction', it would surely qualify as 'metaphysical'; in the sense that it is characteristically the sort of 'conviction', or thesis, that classical metaphysics dealt with. What the sources or reasons for such 'metaphysical convictions' are, is another matter, of explaining metaphysical ideas, or accounting for their genesis (as e.g., Meyerson does with Identity, or as Hegel does, in attempting to explain the logical force of concepts of unity and identity, or as we might do in sociological or psychological terms). Indeed this metaphysical notion of unity is the recurrent theme of classical philosophy, from the Pre-Socratics on, and has its varied expression in as diverse contexts as theology, art and science, and in as widely separate traditions as those we crudely lump as 'Eastern' and 'Western'. And if indeed, the question of what our theories are about is significant for science, then *this* metaphysical 'conviction' (or more modestly, this pervasive and recurrent hypothesis of metaphysics) is basic to scientific understanding. Its varied formal expression in science and philosophy – now as the Heraclitean *Logos*, now as the Parmenidean One, now as the Platonic

unity of Being and Knowing, now as the Principle of Sufficient Reason, now as the 'axiom', or the 'inductive generalization' or the 'postulate' of the Uniformity of Nature, now as the Law of Parsimony, now as the Principle of Least Action, now as simplicity, now as coherence, now as consistency – all attest to its pervasiveness in the history of thought.

But what *is* the thesis that "reality is fundamentally one"? Is it really, as it often appears, a *cosmological* claim, that there is one stuff out of which everything is constituted? Is it only a generalization of the kind of thing which Agassi, for example, takes metaphysical claims to be – (e.g. 'All is water', 'All is force')? I think this trivializes and mistakes the significance of the claim. I would rather say that the cosmological thesis merely expresses a deeper principle, and one which lies at the heart of the Ionian conceptual revolution: namely, that there is one ἀρχή, one *principle* in terms of which any explanation or rational interpretation is possible, and that is, that there is one objective truth, one λόγος, with respect to which our explanations and theories are adequate or inadequate, true or false.

This principle, in effect, formulates the condition of objectivity in science, in the form of a 'conviction', if you like, or in the form of a heuristic principle, the critical examination and articulation of which is the traditional province of metaphysics. It is expressed historically in many ways: e.g. theologically, in the notion that God is one, with its concomitant that truth is one (it being, in effect, what the one God knows it to be); metaphysically, in the notion of one substance, or one essence or form, with which true knowledge is 'identical' (and therefore also 'at one with itself', as in coherence theories and in all the versions of subject-object identity, from Plato through Hegel). One could go on to trace the extraordinary history of this conception. But what are we to make of these esoteric and fantastic formulations? That they are inebriate visions of some mystical sort? Or that they express some unutterable 'ground of Being' of which we are only dimly aware, or which enters into our knowledge tacitly, as Polanyi would have it, or as 'intimation'? Or are we to interpret this as the expression of some deep-rooted desire to return to the symbiosis of the womb, or to the 'whole breast' of infancy; or are we to track it back to our perceptual propensities to close gappy configurations into completed *gestalten*, or to explain it as an '*a priori* condition of the understanding'?

All this, however intriguing, interprets esoterica in a still more esoteric

way. I would propose instead that the metaphysical thesis is only a sys-
tematic and general formulation of a most profound commitment:
namely, that there is a reality which can be known truly, and that it is not
relativized as to time, place or person; that it is the *same* – i.e. 'one' – for
everyone everywhere at every time. The logic of this commitment is also
profound and simple; for only from this starting-point (of a unitary
reality, and truth) does it make sense to talk of the relativization, the par-
tiality, the *criticizability* of our conceptions, our theories and our models.
"Listen not to me but to the *logos*", says Heraclitus, in the first formula-
tion of metaphysics proper. Newton makes it one of the 'Rules of Phi-
losophy' in the *Principia*, but states it in cosmological terms, in terms of
the uniformity of nature. Duhem found this commitment to be an act of
faith, transcending science. In his discussion of our propensity to form
'natural classifications', totally unjustified on his positivist view of science,
he says that "while the physicist is powerless to justify his conviction, he
is nevertheless powerless to rid his reason of it", and says further that it
is fortified by experimental confirmation of predictions.[34] Peirce suggests
this commitment, in effect, as the realist methodological criterion of
scientific questions – those concerning which all scientific investigators are
"fated to agree".

But we may well ask whether, short of voluntarist, or fideist or psycho-
logical 'justifications', the critical scientist is justified in such a deep com-
mitment. Let us grant that the man of common-sense needs this rock to
stand on, if he is to proceed to act rationally. Let us grant that a random
universe, or a throughly relativized one would reduce following any guide
to action to an act of irrational faith or to the existentialist's 'leap'. The
scientist starts, presumably, by questioning common-sense assumptions,
by discovering their vagueness and inadequacy with respect to the fine-
grained domain of those truths which science discovers. Yet, the scientist
too ends up with some commitment, however tentative and however
skeptically qualified, to the proposition that knowledge is cumulative,
that science is an inquiry into truth, and not just blind conjecture. Modes-
ty and skepticism (or merely the fear of professional scandal) may temper
the reach and the sureness of metaphysical claims, for the Absolute,
damned by James, remains shy. Yet the claim to approximate truth en-
tails a claim to that truth which is approximated, and which is itself
therefore, *not* approximate. It is this metaphysical commitment, to objec-

tivity, or to objective truth, which bears the weight of the heuristic function of metaphysics in science.

What may justify such a commitment? Or is justification precisely what this commitment does not lend itself to? I would like to propose what I think does justify such a commitment, and what accounts thereby for the heuristic function of metaphysics in science. Clearly, even such a root-commitment requires justification, on my view; like every other thesis in science and philosophy, it *is* articulable and may be subject to criticism. For even the view that this commitment is 'animal faith' is itself a philosophical thesis subject to criticism, and does not close the door on argument, but merely opens another of many doors.

My thesis is this:

(1) That metaphysics formulates in systematic and explicit form the deepest and most pervasive features of common sense. And that common sense is that set of ready-to-hand, and largely tacit conceptual formulations (in practice: operative guides, maxims, rules, beliefs) which have withstood the most severe and persistent trial, in the common practice of a society of men, in the developments of their cultures, and in the trans-cultural contexts in which one may speak of human universals. I take common sense, then, to be the natural and social product of an evolution extending at least as far back as human discourse, in which selection has proceeded by the test of common practice and experience, in the experimental crucible of human, social existence. There is no claim here that common sense from one culture to the next is the same, and I would suspect that it changes with variation in experience. Nor is common sense an explicit system, so that it is possible that common-sense concepts are inconsistent, incomplete, incompatible with each other because not subjected to the tests of rational criticism which require systemic articulation. But where common-sense concepts persist, then the matrix of experience in which they persist is itself persistent, and may come to be taken as 'permanent' or 'universal'. But the critique of whatever sense is common is the birth of philosophical reflection, whether as ethical criticism of beliefs and practices, or as cosmological criticism of myths. The hallmark of criticism is always that it attacks the accepted 'permanencies' and 'universals', the dogmas which an *uncritical* common sense erects finally into systems of articulate belief and practice. These, the laws, rituals, maxims, myths, rules of practice, taboos, go beyond the tacit content of common

sense, embody it in *criticizable* form, but not yet in critical form. It is from this cluster of articulations that metaphysics arises. At the same time, because metaphysics is systematic and explicit in its formulation, it both reveals the incompatibilities and inconsistencies in common sense plainly, and also introduces the critique of common sense. But this would then seem to make common sense the first victim of criticism, and that it persists still would seem to be due to a perversity immune to criticism. Here, perhaps, the dynamics of change in common sense come into play: for with the birth of criticism, the matrix of common experience itself changes, for now ordinary human practice and experience includes not simply the rules of thumb acquired in sheer practice, nor simply the structural propensities of the organism which it inherits, but also the articulateness of critical discourse. Our own common sense is the product, not simply of some brute practice, but of a common inheritance, and a roughly common education which brings to bear, in whatever tacit and distorted forms, the conceptual influence of philosophical and scientific theories, and of religious, political, and aesthetic theories, passions, styles and dogmas, all of which are as much a part of the matrix of our common sense as are the solid objects which we cannot walk through, and the other minds with which we communicate. It is precisely 'permanencies' and 'universals' in this matrix that metaphysics attempts first to formulate in systematic ways, then to criticize. But this twin function already generates an ineradicable tension at the heart of the metaphysical enterprise: what often is characterized as the dual function of philosophy: speculation, and critique. Metaphysics is the critique of what is taken to be permanent and universal, and is at the same time the formulation of universal claims. The central concern of metaphysics with what is universalizable, generalizable, with unification of diversity under some principle, is so pervasive, in the history of philosophy, that the sense of its origin, and the direction of its application are often lost; in the abstracted, or bowdlerized, or simply sophomoric versions one gets (either in bad metaphysics, or in the mythical metaphysics that enters the text-books and histories, or in the scarecrow images of early positivism), metaphysics is some wildly irresponsible, or at best merely literary-poetic speculation about some grim, or gorgeous, or ethereal or substantial Absolute. The pity is that such versions permit neither the proper critique nor appreciation of metaphysics. In more serious versions, and with the sophistication of some

knowledge of historical and social and technological contexts, the role
of metaphysics in the articulation and critique of the most pervasive
features of common sense becomes clearer. And then it becomes clear too
that the philosophical critique which metaphysics introduces becomes
part of the testing of common-sense concepts, and operates now in that
process of selection in which our present common sense has evolved.

It needs little elaboration to make explicit the corollary thesis here:
Science, as the continuation of the critical and constructive functions of
metaphysics itself, becomes the test of common-sense beliefs and prac-
tices; but only when the test itself becomes explicit, and is rationally
related to common sense; that is, when science does not become so
esoteric, nor common sense so degraded and stupid that the testing of
one by the other is simply unfulfillable. The contemporary task of phi-
losophy of science is in large part exactly that of confronting our present
common sense with the critical test of science, and therefore, with art-
iculating the content of scientific theory in such a way that such a test
becomes feasible. That this entails the critique of scientific concepts as
requisite to such an articulation is a simple consequence of such a formu-
lation. In this sense, Philipp Frank was profoundly right in his charac-
terization of the philosophy of science as a bridge between the sciences
and the humanities (though what is then required is an adequate charac-
terization of the humanities. I can only refer to Norman Rudich's per-
spicuous analysis of some of these problems in the previous volume of
these *Studies*).[35]

(2) The structures of metaphysical systems are derived from the struc-
tures within which common-sense concepts are formulated, and ulti-
mately, from the structures of common action and experience of a human
society. I would further suggest that a major mode of access to these
structures is to study them, in their articulate and in their hidden form,
in language. Just as we may reconstruct the mode of life of an organism
from the evidence of its structural (anatomical and physiological) make-
up, I suggest that we analogously reconstruct the common modes of
action and experience of the human organism from the evidence, and the
impress of these modes in language. I mean this in the sense of so-called
deep-structure of language, which I would take to be philosophically
interesting because it is more than the *formal* structure of language (though
one may study it effectively as formal or theoretical grammar in the

Chomskyan sense of grammar as a theory of the linguistic competence of
a speaker). The very conditions of language and of communication
(beyond primitive gestural and expressive communication) are those
which metaphysics exhibits in a relatively abstract form: systematic struc-
ture, reference, and abstraction. One may argue that in order of logical
priority, the elements of language (whether one takes these as morphemes,
or as sentences) precede the use of language to tell a story. But I would
argue that the story-telling *function* of language is a necessary condition
for its evolution as an artifact, and that even the primitiveness of osten-
sion and denotation, on the usual accounts, is compromised by the prior-
ity of this function. In its substantive content, as a structure – apart from
a purely formal or syntactic mapping or theory of such structure – lan-
guage embodies, and helps to shape, the 'hypotheses' of common sense.
I would further suspect that the most pervasive and persistent ones, from
the point of view of primary experience, become embedded in the very
structure of language itself; which is to say that they yield very little to
local variation, the artifact acquiring a certain autonomy once formed.
(Certainly, if the 'innatists' in linguistics are right, then the plasticity is
even less, for the consequence of a strong form of the innateness hypoth-
esis is that language has a genetic base which is species-specific in
human beings. See the symposium on 'The Innateness Hypothesis', in
this volume.)

(3) Metaphysics, in introducing the conceptual model as an articulate,
explicit object of criticism, represents the way back to a critique of the
foundations of our understanding. It makes of the conceptual model not
only a practical instrument of the understanding, in *use*, but an object
of it, in *inquiry*. Thus, it introduces conceptual criticism by making the
object of such criticism fully and rigorously explicit; for in the absence
of such an explicit theoretical object, explicit criticism is itself impossible.
In this dual function of articulating and criticizing conceptual models,
metaphysics serves as a heuristic for the understanding itself: it is *self-
critical* and examines its own foundations. (Obviously, it is not every
metaphysician who does this for his own system; but the sense of the
metaphysical tradition is that if he does not go at his own system, some-
one else will.) It is only when the tradition is ignored, when a metaphysics
holds itself immune from critique that it degenerates into *non*-metaphys-
ical ritual and dogma. That is why the study of the history of metaphysics

– of the alternative and incompatible systems, and of the sustained and rigorous critique of these systems – is itself a heuristic for the understanding. Sometimes, in moments of greatness, metaphysics is self-critical in the person of a single metaphysician – e.g. Plato, in his critique of the Theory of Forms in the *Parmenides* – but this is rare and atypical. The requirement for a community of criticism, presaging the scientific community, is central therefore, as the condition for rationality.

(4) Finally, metaphysics is heuristic for science in providing the root-model for scientific understanding. As an exercise in self-critical theory construction, it not only (incidentally) provides anticipatory models for science, but perhaps more importantly, formulates the conditions for conceptual representation in a model, as conditions for understanding. We do not come to understand something simply because it is presented in some conceptual model, but because the conceptual model, as a form, is consonant with the very primary experience of the understanding itself, with the story and with the structures of language which embody the story-telling function. Now science is not story-telling, in any simple sense; thus, to take metaphysics as a way back from scientific theories and models to their 'recounting' in ordinary language or in terms of common sense is to miss the point; for this has not anything to do with the *scientific* understanding, but only with the *popularization* of science. Thus, Frank, some years back, confused the roots of metaphysics in common sense with a notion of metaphysics as the common-sense vulgarization of science.[36]

My argument has been, rather, that even at the most advanced levels of scientific theory, the formation of concepts and of conceptual models bears the impress of that form which story-telling exhibits in its concrete imagery, which metaphysics develops into rigorously abstract systems, and which, I have claimed, goes back to the primary experience of coming to understand. Nor is this an argument in terms of familiarity and unfamiliarity – like those arguments which hold scientific explanation to be the representation of the unfamiliar under the form of the familiar. For this is to mistake the superficial or surface-structure of the story for its deep-structure. And although what I propose may sound like some Kantian *a priori forms of the understanding*, I take these forms to be generated and evolved in the course of social human practice, as language itself is. In the sense, metaphysics is not a heuristic for the understanding

of science by non-scientists; rather, it is the bridge between advanced and intelligent scientific *practice*, and rational scientific *understanding* (in the special senses in which I defined 'intelligent' and 'rational' earlier).

My concern here derives in part from the patent fact that some of our most advanced scientific practice – and here I do not mean simply 'practical' or 'applied' science or engineering or sheer laboratory activity, but the broader practice of science which includes theoretical science as well – is vitiated by an irrationalism: one in which leading scientists have quite explicitly given up the task of scientific understanding, or have conflated it with or reduced it to the predic⁺ive and instrumental aspects of science.

One may appreciate, in its historical context, that clearing of the ontological overgrowth which positivist and operationist criticism undertook. It brought to bear the full force of contemporary advances in methodology and logic, and sharply posed the epistemological questions which arose with the new physics. But, as I have tried to show, both the strong and the modified forms of its reduction program have failed; and for profound philosophical reasons. It sought to make do with a methodological algorism in place of an ontological heuristic. But for the reasons I have adduced, I believe the scientific understanding requires more than a heuristic for successful practice; it requires a critical theory of reality. The abandonment of such theory, or the assignment of it to the limbo of *non*-philosophical 'external' questions opens the door to arational and irrational reliance on intuition, or faith or some atheoretical sheer 'practice'. It also leads to the sharpest breach between science and the common understanding, refusing to educate critically at the point where such education is most essential. This is perhaps the worst damage that the originally critical heritage of positivism has done to science.

So much for the study of metaphysics as heuristic for science. What about actually *doing* metaphysics? Shall the scientist abandon his science to 'become' a metaphysician, in the interests of 'understanding'? Or shall a small, hardy band of philosophers undertake to render science their services as metaphysicians, bearing the gift of 'understanding' to the blind, benighted practitioners? Both alternatives misconstrue the activity. I think that within each domain of scientific practice, the conceptually self-conscious and self-critical scientist *is* doing metaphysics and has been doing it all along. He is, most often, *métaphysicien malgré lui*, in proposing and examining conceptual models in his domain. Whether a metaphysics

of broader scope is possible, whether that theoretical and conceptual unification of the sciences is more than a hope, I do not know. The working hypothesis of the unity of science has its ontological correlate in the unity of nature, and here as elsewhere methodology finds its correlate in ontology. But the testing of this hypothesis requires not simply philosophical, but scientific work – or better still, philosophically sophisticated scientific work. One may say that such a program underlies our best hopes for science, as the theoretically critical commitment to some fundamental reality underlies our best hopes for truth. Whether new sweeping metaphysical systems of universal scope remain to be constructed, and whether they will be crucially different from classical ones I really do not know; but it is likely that they will be attempted; and they should be attempted only with the clear-headed and hard-headed awareness of the mutual relevance of scientific theory and discovery on the one hand, and critical metaphysics on the other. There has been only one such grand design in this century, whose scope, rigor and depth earn for it such a title, as far as I can tell; and whose difficulty has left it, for the present, without significant heuristic force for science. Perhaps it remains to be rediscovered.[37]

Boston University

NOTES

[1] In Popper's case, 'recent' needs qualification, since much of what I will discuss is in the 1934 *Logic of Scientific Discovery*. Nevertheless, it is a component of recent discussion, both in the sense that its theses are elaborated and expanded in Popper's more recent essays, and in the sense that the earlier work came under full consideration only with its translation and publication (in revised form) in English, in 1959.
[2] Carl Hempel, 'Carnap and the Philosophy of Science', in *The Philosophy of Rudolf Carnap* (ed. P. A. Schilpp), LaSalle, Ill., 1963, p. 707.
[3] Y. Bar-Hillel, 'Carnap's Logical Syntax of Language', *Op. Cit.*, p. 537.
[4] *Ibid.*, p. 536.
[5] Willard V. O. Quine, 'Carnap and Logical Truth', in *Op. Cit.*, p. 405.
[6] William Craig, 'On Axiomatizability within a System', *Journal of Symbolic Logic* 43

(1953) 30–32; and 'Replacement of Auxiliary Expressions', *Philosophical Review* **65** (1956) 38–55.

[7] Carl Hempel, *Op. Cit.*, p. 699. See also Scheffler's discussion on Craigean and Ramseyan elimination, *Anatomy of Inquiry*, New York 1963, pp. 193–222.

[8] Macquorn Rankine, 'Outlines of the Science of Energetics', *Miscellaneous Scientific Papers*, 1855, p. 209; cited in P. Duhem, *The Aim and Structure of Physical Theory*, Princeton, N. J., 1954 (Athenaeum, 1962), p. 53.

[9] William Thomson, Lord Kelvin, *Lectures on Molecular Dynamics and the Wave Theory of Light*, Baltimore 1884, pp. 131–2.

[10] Pierre Duhem, *The Aim and Structure of Physical Theory*, pp. 70–71.

[11] Karl Popper, 'The Demarcation Between Science and Metaphysics', in *Philosophy of Rudolf Carnap*, p. 198; see also *Conjectures and Refutations*, New York 1962, pp. 267ff.

[12] Karl Popper, *The Logic of Scientific Discovery*, London 1959, p. 43.

[13] Karl Popper, 'On the Status of Science and of Metaphysics', in *Conjectures and Refutations*, p. 196.

[14] Popper, 'Demarcation . . .', *Op. Cit.*, pp. 278–279; in *Carnap*, pp. 211–212.

[15] Joseph Agassi, 'The Nature of Scientific Problems and their Roots in Metaphysics' in *The Critical Approach* (ed. M. Bunge), Glencoe, Ill., 1964, pp. 189–211.

[16] Joseph Agassi, *Op. Cit.*, pp. 191–193. Imre Lakatos also addresses himself to this thesis in 'Demarcation Criterion and Scientific Research Programs', in *Problems in the Philosophy of Science* (ed. by I. Lakatos and A. Musgrave), Amsterdam 1967.

[17] In the quasi-Popperian sense of 'part of science' which Agassi holds to in this paper.

[18] *Ibid.*

[19] Popper, *Conjectures* . . ., p. 55–56.

[20] J. Agassi, 'The Confusion between Physics and Metaphysics in the Standard Histories of Science', in *Proceedings, XI International Congress in the History of Science*, Ithaca 1964, p. 233.

[21] J. Agassi, 'The Nature of Scientific Problems. . .', *Op. Cit.*, p. 198.

[22] George Berry has suggested to me, on this point, that since from '$(M \to T) \cdot (T \to P)$', it follows that '$(M \to P)$', by the transitivity of implication, then, if the scientific status of 'T' lies in its making a logically true conditional out of '$(T \to P)$', then, if '$(M \to T)$' is logically true, 'M' must also enjoy this same status. Agassi's point, however, is that in the case of 'M' such that '$(\bar{M} \to T)$' is true, then 'M' is still empirically refutable. (All this assumes that on even a qualified interpretation of '$\to$' in '$(M \to T)$' or '$(M \to P)$', *Modus Ponens* remains our rule of inference.)

[23] Thomas Kuhn, *The Structure of Scientific Revolutions* (International Encyclopedia of Unified Science, Vol. II, No. 2), Chicago 1962, p. 41.

[24] *Ibid.*, p. 4.

[25] *Ibid.*, pp. 23–24.

[26] *Ibid.*, p. 110.

[27] T. Kuhn, *Proceedings. . ., Op. Cit.*, p. 248.

[28] J. Agassi, 'The Nature of Scientific Problems. . .', *Op. Cit.*, p. 210.

[29] M. Polanyi, 'Discussion. . .', in *Scientific Change* (ed. A. C. Crombie), New York 1963, p. 375.

[30] C. S. Peirce, 'Notes on Scientific Philosophy'.

[31] It is clear that there is understanding in science which is *ad hoc*, or which falls short of theoretical understanding of the 'grand system' sort. Of course, there are degrees of understanding, from, e.g., understanding some simple apparatus, or some delimited experimental result, to understanding a 'grand theory'. I would argue that all of these

require some kind of theoretical context, if what we are talking about is scientific understanding, and not simply some technical skill.

[32] Mary Hesse, 'Models and Matter', in *Quanta and Reality*, Cleveland 1964, p. 56.

[33] *Ibid.*, p. 56.

[34] P. Duhem, *Op. Cit.*, p. 27; Cf. also pp. 24ff., 293ff.

[35] Norman Rudich, 'The Dialectics of Poesis: Literature as a Mode of Cognition', in *Boston Studies in the Philosophy of Science*, Vol. II (ed. R. S. Cohen and M. W. Wartofsky), New York 1965, pp. 343–400.

[36] Philipp Frank, 'The Place of Logic and Metaphysics in the Advancement of Modern Science', *Philosophy of Science* **40** (1948) 275–286.

[37] The reference is, of course, to Whitehead. For a recent treatment of Whitehead's relevance to contemporary physics, see Abner Shimony, 'Quantum Physics and the Philosophy of Whitehead', in *Boston Studies, Op. Cit.*, pp. 307ff. and the comments by J. M. Burgers.

MATTER, ACTION AND INTERACTION*

[1973]

ABSTRACT. In this paper, I will argue that a materialist ontology is an important heuristic for scientific theory, and that a classical problem in natural philosophy – whether matter is self-active or is inert – has its contemporary counterpart in modern physics, especially in microphysics. From this, there follow epistemological and methodological consequences concerning scientific inquiry and practice. In order to set the problem of my paper, I introduce as background the contemporary discussion of the relation of metaphysics to science. The *philosophical* problem of the paper, however, does not concern the formal or methodological issues in this debate, but rather how these issues are resolved in a concrete case, concerning matter, action, and interaction.

INTRODUCTION

There has been much discussion in recent years, first about the 'demarcation' between metaphysics and science, and then about the relevance of metaphysics to science, and the lack of a clear demarcation. Several traditions, both in science and in philosophy of science clash here. The earliest impetus to the 'de-metaphysication' of science, at least in modern philosophy, came from the positivist attempt to free science from 'speculative metaphysics', and in particular, from idealist metaphysics. Thus, Mach, for example, introduced a version of empiricism, in which 'metaphysical' questions were reduced to empirical questions, or when not reducible, were eliminated. John Stuart Mill, earlier, had effected such a 'reduction' in his definition of matter as 'the permanent possibility of sensation'. (Though 'permanent possibilities' seem no less 'metaphysical' than does 'matter,' it appeared to delimit Sir William Hamilton's 'unknowable', and to establish what has been called 'matter' as that which is sensible.) Later positivism eliminated all such 'metaphysical' talk, first by a rigorous phenomenalism (e.g. in Rudolf Carnap's *Logische Aufbau*) and then, by a linguistic reductionism, in which the choice of language-frameworks for talking about scientific questions became a pragmatic question of 'expediency', and all so-called 'external' questions (e.g. concerning what exists) were placed outside philosophy of science proper as 'practical' questions to be resolved on pragmatic grounds. Sir Karl Popper's contribution to the elimination of metaphysics from science came by a somewhat different route, namely

90

that of 'falsificationism': Science was the domain, according to Popper, of those statements which could be falsified by so-called 'basic statements' (or their consequences); and basic statements, in turn, were acceptable on grounds of empirical testing procedures, (though Popper, like Carnap, ultimately based the *acceptability* of these basic statements on social practice and expediency). Metaphysics, by contrast, was (according to Popper) the domain of the unfalsifiable; and therefore, being empirically untestable in its claims, it stood clearly demarcated from science.

The philosophical problem with such 'reductionist' or 'eliminationist' or 'demarcationist' views is that they 'eliminated' not only speculative idealist metaphysics, but *all* ontological claims, including materialist ones. Carnap's 'physicalism' had nothing to do with the *existence* of the physical world, but only with the choice of certain primitive terms, or certain *Protokolsätze* in a language; or the choice of a particular 'physicalistic' linguistic framework, on grounds of its utility or expediency. Popper's 'realism', which claimed for science the objective search for truth as the goal of inquiry, likewise eschewed 'metaphysical' talk about 'reality', and thus ends up as an epistemological 'realism' without 'reality'. Popper's argument is at least ontological in some degree, since the very basis for discovering falsehood entails that something is *not* the case. But in his long flight away from Hegel, and especially from Marx, Popper retained only the negative element of the dialectic, the view of the fallibility of knowledge and the relativity of our truth claims, but did not permit himself the positive consequence, that to be fallible, or relative, there must be that with respect to which our knowledge *is* fallible or relative – i.e. objective reality as an *object* of knowledge.

There have been counterarguments to this anti-metaphysical view, in recent years. Thus, in a historical-descriptive way, without yet getting to the philosophical questions, Thomas Kuhn[1] argued that, in scientific practice, metaphysical ideas played a role in the characterization and formulation of problems. So, too, Joseph Agassi, after demarcating 'good' from 'bad' (irrationalist) metaphysics, has argued that metaphysical ideas are embedded in major scientific theories, and showed how the uses of the term 'metaphysics' are often equivocal and lead to confusions.[2] I argued, some years ago, that neither Popper, nor Kuhn nor Agassi went far enough in their assessments of the role of metaphysics, or in their anti-positivism; and that metaphysics has played

a heuristic role in scientific discovery.[3] And more recently, Lakatos has argued, in his paper on 'research programs', for the 'negative' and 'positive' heuristic role of metaphysics. So too, Feyerabend has argued that dialectical materialism "as it appears in Engels, Lenin, Trotsky," *synthesizes* the two co-present aspects of scientific progress: tenacity and proliferation.[4] If *materialism* is clearly a 'metaphysical' view, dialectical materialism certainly is as well! And so, at least methodologically, Feyerabend argues for the positive role of some metaphysics in science.

I do not want to enter here into the discussion on metaphysics and science only in a formal or methodological way once again. Rather, let me state what I think the *philosophical* issue is, in order to deal with it in the more concrete context of actual philosophical problems in physical science. That is, I do not want to talk programmatically, *once again*, on the relation of metaphysics to science *in general*, but rather to discuss a *specific* philosophical-scientific issue in which this relation can be demonstrated, namely, on the concepts of *matter, action*, and *interaction* in natural science; and on the epistemological and methodological consequences of one or another 'metaphysical' or 'ontological' view. (I will use the terms 'metaphysical' or 'ontological' interchangeably, to denote those statements or views which make claims about the nature of reality. This is a simple and classical view, and I urge everyone to adopt it once again, now that positivism has lost its power to terrify philosophers into silence and embarrassment over 'reality'.)

What, then, is the philosophical problem? Put very simply, it is: Is matter self-active or not? Now one may argue (1) that this is not a 'philosophical' problem at all. Either it is a pseudo-problem, (i.e. a 'metaphysical' problem without solution); or (2) that it is a question of what view it is more 'expedient' or 'useful' for a scientist to hold or accept, and therefore a practical (and pragmatic) question for scientists, and not for philosophers of science. Or, again (3) one may argue that the problem has been resolved in modern science, since matter and energy are simply two alternative modes of an *admittedly* 'self-active' physical reality, and 'inert' matter has been abandoned, or at least quarantined to the domain of an older classical physics. Or one may argue that, though it is a 'philosophical' problem, the concept of *matter* as a general category is outside the domain of science, and that scientists deal only with specific and concrete features of the physical universe; i.e. that the demarcation between science and metaphysics holds.

I will not argue against the first two views here (i.e. the positivist

'elimination' of the problem as a *Scheinproblem*, or the pragmatist-instrumentalist 'resolution' of the problem as a (non-ontological) question of practical choice among alternatives.) Rather, I will address only the third argument, that modern physics already accepts the self-activity of matter, and that there is no longer a problem; or at least, no problem *within* science.

In order to do this, however, let me first briefly state the older, traditional formulations, and then show how they carry over, and are transformed in modern physics.

<p style="text-align:center">I</p>

The classical characterization of matter as *inert* formed the philosophical basis of much of 17th Century physical science. At the same time, the counter-tradition viewed matter as self-active or in self-motion. In this paper, I propose the thesis that these two contradictory metaphysical views still influence contemporary physics and philosophy of science, though in entirely novel ways; and that they have consequences for epistemology which roughly demarcate two views of scientific knowledge.

First, what is at issue, methodologically, in the counterposition of these two views? Let me restate the traditional views in a modern context in order to answer this question. On the view that matter is inert, or that it has no self-activity, we may translate this into the view that change of state, either at the atomic or the sub-atomic level is always initiated by an interference with a state from without – i.e. from outside the system or subsystem under consideration. There are, on this view, no 'spontaneous' atomic or sub-atomic events, whose genesis is *within* the system itself. Methodologically, this is one variant of strict determinism, since for every change of state of a system, we would then look for its cause *outside* the system, as a change in energy, charge, etc. Thus, for example, if in a bubble-chamber photograph, the trajectory of a particle changed abruptly (e.g. in index of curvature, or in bifurcation, etc.), we would be methodologically constrained to hypothesize an interference from 'outside' that particle itself – e.g. a collision – to explain the change. Moreover, on conservation grounds, the change in trajectory, as a phenomenon, would have to be explained by an event in which conservation of momentum, or charge, or energy would be systemically preserved. Thus, if one were to take a *classical* view of elementary

particles, one would expect that they would behave as Newtonian particles, i.e. any change in motion would be caused by a change in the force acting upon that particle. On one reading at least, such acting forces would be external forces.

There is, however, an alternative tradition, in which a change of state is brought about not extra-systemically, but intra-systemically. So, for example, in classical LaPlacean determinism, if all the positions and momenta of all the particles *in a closed system* were known, and all the laws of their motion were known, one could predict (or retrodict) the state of the system at any *other* time, given the state of the system at any given time. Further, this epistemological consequence (given the premise of omniscience) is based on the ontological premise that such changes of state through time are strictly determined. But this determinism has nothing to do with 'external' forces acting upon the system, but only with the relations among *intra*-systemic forces; i.e., for the system *as a whole*, the relations among its constituents, or its 'internal' relations totally determine the dynamics of the system. But such intra-systemic changes account for *all* temporal events in the system only if there are no external forces operating upon the system from without, i.e. only if the system is isolated and closed. LaPlace's universe was, of course, such an isolated system, since it was *the* universe, and this monism admitted nothing and required nothing beyond this. That is to say, it admitted no extra-universal forces. He was thus able to reply to Napoleon's query as to whether he believed in God's existence, with the statement "Je n'ai pas besoin de cette hypothèse."

This monism defines a determinism which proceeds by the 'necessary' internal relations of a system. Such a monism is that of Spinoza's *substance*, which, as he says, is *causasui*. Its activity is self-activity.

To return to our bubble chamber: we now observe a trajectory which swerves, or bifurcates, or ends abruptly. We may, of course, hypothesize, for example, a collision or an absorption – i.e. an interaction between the observed particle and something external to it. Or we may hypothesize that the particle is a system with internal relations – i.e. with internal structure, and that therefore, its trajectory-changes are caused by its internal changes; and that they are, so to speak, *causa sui*. What appear to be 'spontaneous' or 'random' events are in fact, on this hypothesis, functions of internal causality, to be explained, presumably, by a theory of the internal structure of the particle. At once the question arises, whether such a particle is 'elementary', or whether, if

it has a constituent structure, this structure itself may not reveal 'parts', since the internal relations of such a 'system', as the particle is now hypothesized to be, must be relations among constituent 'parts', *more elementary* than the 'elementary' particle itself.

We are in something of a dilemma. For it certainly is the case that such elementary particles are not *isolated* systems, even if they are systems; and therefore it is conceivable that changes of state of such 'particle-systems' are caused by externally acting forces. But insofar as they *are* systems, such changes of state may be caused by the systemic internal relations themselves. In the first case, where the spatial and temporal contexts of interactions are so small, and the energies are of quantum order, macroscopic observation may not, in principle, investigate the process directly. But then, if the macroscopic evidence yields the conclusion of a change of state, and there is no 'classical' evidence of an interaction – e.g. a collision – the 'classical' determinism in terms of external forces acting upon the particles may be preserved by hypothesizing as yet unknown variables *within* the system, or beyond this, hypothesizing so-called 'hidden variables' which are forbidden by the established formalism, as the appropriate 'external forces'. The dilemma which presents itself, where the conditions of quantum-theoretical observation constrain any more or less 'direct' investigations, is then between hypothesizing 'internal' (i.e. structural) vs. 'external' (i.e. interacting) forces to account for changes of state.

There is a way out of the dilemma, of course, since the dilemma presupposes that the changes of state are caused either by internally-acting or externally acting forces, and that these two accounts are mutually exclusive. On the 'internal' account, the particle is self-active, in *changes* of state; on the 'external' account, the particle is *inert* in such changes. But suppose these are not mutually exclusive, and there are *two* components of change of state possible, and that both may act at once, so that a given change of state is a composite event, made up of a combination of internally and externally-acting forces, and that even interactions between particles involve such simultaneous activity. (In Leibniz's terms, even the 'passivity' or mode of reception of an action by a particle would itself be a function of its internal structure, and a kind of activity.)

But does this really resolve the dilemma? Or does it rather complicate it? Since, now, the heuristic character of the model of combined internal-external activity leaves it open to a choice as to which of the

forces in a given case one should attribute the change of state, or to what extent. We have, in effect, not only interacting, internally complex particles, but at the epistemological level interacting *models* – the inert or passive, and the self-active – and we are in a situation in which the choice is empirically or experimentally underdetermined. We are, so to speak, free to choose various alternative modes of the combination of forces as hypotheses. And in fact, that is what has been done theoretically. The difficulties come in devising experimentally testable consequences of the alternative theories, e.g. of hidden variables, of internal structure, etc.

But all of the foregoing discussion is based on a classical and more or less pictorial model of *particles, trajectories, interactions*, and this model, in turn, depends on the more general (metaphysical) model of a discrete or particulate universe. The question becomes more complex when this model itself is confronted with theoretical alternatives: either in classical field theories or in classical quantum theory. In the first, changes of state are changes in a field, or in field strengths; in the second, changes of state are no longer associated with particulate events, but are rather mass-events, concerning ensembles of particles. Nor are these two alternatives the same: the classical field-description is not simply a statistical description in terms of 'mass-events' involving ensembles of particles. However, these two alternatives have yielded different 'mixed' models. Thus, for example, in classical field-theory there is the interaction of particles and fields; and in quantum field-theory, the 'interaction' may be spoken of in terms of 'particles guided by fields'. But even here, in quantum field-theory, the duality of particles and waves reappears in the so-far unresolvable differences between Bosons and Fermions, i.e. those field phenomena describable by Bose-Einstein statistics and those describable by Fermi-Dirac statistics. In the classical limit, the Bosons tend toward classical fields, while the Fermions tend toward classical particles, reintroducing the duality of particle and field within field theory.[5]

In a more speculative *pure* or *unified* field theory, on the other hand, there are no 'particles' at all, but only variations in field-characteristics. Yet there is an *approximation*, in field-theoretical terms, to the concepts 'particle', 'trajectory', 'interaction', as these are understood *as* local variations in a field. The field however, is a structure, and has 'internal' relations, or self-activity. Thus, Descartes hypothesized 'vortices' in a field-theoretical account of matter as a plenum, and this idea has had an

analogous contemporary development as well. Now if, on a statistical or probability account, the very mode of existence of a particle depends on its interactions, so that it may be 'smeared' all over a space, in one state, and 'contract' in another – e.g. in an interaction, in which it leaves a macro-trace e.g. in a scintillation counter, or on a screen, or in a bubble-chamber – then the account of 'self-activity' vs. action by external forces in a change of state becomes more complex than in the classical picture. The wave-particle duality, the interaction of fields etc. all become involved in the account, and coherent models, which retain classical-particle interpretations or translations, become impossible. In effect, so-called 'anomalies' appear, if one chooses to hold on to classical models of location, trajectory, interaction, etc., but are resolved if one adopts a complementarity view, in which the classical requirement for simultaneous values of the canonically conjugate variables (e.g., position, momentum) is given up.

One may, of course, conceive of treating fields as if they were large and complex entities with internal structure, which have boundaries and which therefore have an 'inside' and an 'outside'. As field strengths may tend towards zero, – e.g. in vector-addition of positive and negative field forces – the field still 'exists' in such a region, since the field equations are still continuous. But one may speak of 'field-boundaries' in this limit sense. If one gives up the notion of definite 'internal-external' boundaries altogether, one may devolve upon a unified-field model of differential field-strengths through all of space, i.e. a model of concentrations and diffusions which goes back as far as Anaximenes. But if there were 'holes' in such a field, i.e. field discontinuities at the limit, we would once again have (negative) particulateness, i.e. regions of field discontinuity (where the field-equations break down) which may be taken as analogous to the 'particles' in classical space. So too, if field-strengths tended towards infinity, then, at the limit, there would be concentrations of a particle-like nature. As field-strengths varied between these limits, there would of course be approximations to such discontinuities, but no actual discontinuities: there would be no 'duality', but rather very steep gradients in field-strengths.

But on such a model of a pure unified field theory, the very notion of the 'inertness' of matter would be difficult to formulate. (Prof. Stachel has suggested to me that the distinction between linear and non-linear field theories may however preserve this notion of relative 'inertness', in the linear case, for only the non-linear fields are described as

'self-interacting'.) On a pure field theory, so too would the notion of
'externally-acting forces' be difficult to formulate. In a 'mixed' field
theory, of course, where there are particle-field interactions, one could
then speak of the relative 'inertness' of particles 'guided' by fields; but
even here, we may speak of a system in which these are *inter*actions, i.e.
internal relations within a closed system of relatively differentiated
constituents. But of course, to conceive of a 'closed' system, in this
sense, is to eliminate 'externally acting forces' by definition. If a field
could be defined, i.e. if its 'boundaries' could be established, then, of
course, 'one' field may be said to interact with an 'other', which is
'external' to it. But without field discontinuities such 'boundaries' are
relative and not definite[6], and therefore between 'different' fields there is
no difference except as local variations in a unified and unbounded field.
Thus the propagation of waves in such a field would simply yield peaks
and valleys of field strength, and so-called action-at-a-distance would
simply become a field-pattern. If differential local characteristics could
then be said to constitute the 'internal structure' of a field, then
'interactions' in a field would simply become internal relations within
such a structure, without any conceivable 'externally-acting forces'.

So too on the statistical-ensemble view. For then, the ensemble or the
system as a whole would be only *relatively isolable*, since the Ψ-function
would describe the distributions in an interaction with macrosopic
observers or measuring instruments. The *localizability* of the 'particles'
would not be a meaningful concept unless the activities of preparation
of a state and measurement took place. In principle, the notion of
'externally-acting forces' could be reconstrued *within* the system of
'observer-observed' so that only the intra-systemic or internal relations
of such a system need be taken into account. Yet, there is a certain
'isolation' presupposed by such a system, and it is this (relative or
constructed) isolation which, so to speak, yields the Ψ-function
predictions for the probabilistic results in measurement.

Thus, it would appear that the distinction between 'inner' and 'outer'
vanishes either on a closed field-theoretical model or on a closed
statistical-ensemble model, and is retained only for open systems. More
generally, the notion of 'inertness' depends on a notion of
'externally-acting forces', and where our theoretical models are closed
systems, either as fields, or as systems of closed interactions on the
statistical-ensemble view, the notion of 'inertness' becomes relativized to
elements within an intrasystemic relation, at best; and perhaps not even

then, except in a perspectival sense. That is to say, in such contexts, 'inertness' is only a phenomenal appearance relativized to observers, and to the theories under which such observation takes place.

But this leads to a strange result. The notion of 'externally-acting forces' upon an 'inert' matter would then seem to be limited to a certain anthropomorphism – i.e. an imposition upon the world-picture of the experience of the macro-world, and in particular, the experience of human action, and the separation of the actor from the object of his action. Now, on a fully monist view, this interaction itself may be seen as 'intra-systemic', i.e. as internal relations within a single material universe. 'External relations' themselves may be understood, then, in the sense of relations between *relatively* differentiated constituents of matter, but not as relations between absolutely differentiated entities. At the macro-level, the relative differentiation plays a major role in the self-activity of the system of the material universe as a whole. But so called 'inertness' – i.e. lack of self-activity – can then be seen as a hypostatization of a *human* mode of activity, – i.e. action upon a 'dead' or 'inert' nature by an active or self-active agent, man. The 'inertness' of nature, at the micro-level, as we have seen, depends, conceptually, upon the notion that the systems dealt with are open systems; and is already qualified by the notion that internal structure of the particles affects the way in which so-called externally-acting forces are received. That is to say, *every action is an interaction*, and *absolute* inertness has no place in the picture. But if we go still one step further, and construe even this relative particularity *either* in statistical-ensemble or field-theoretical accounts of matter, then this *relative* inertness is still further relativized as *every interaction is seen as an internal relation within the system of matter, as a whole*.

II

There are several traditional ways of escaping this consequence. First, one may take the whole system of matter (including man and his activity as a material being) as inert, and postulate an immaterial source of activity *beyond* the system. This is the traditional argument for a First Cause or a Prime Mover, derived from Greek metaphysics and theologized in Christian cosmology. It is the basis for dualisms of every sort, including the mind-body dualism, since human *self-activity* is there ascribed to the mind, but denied to the body, as in Descartes. Second,

one may escape the requirement of *causal* explanation of the changes of state, at the micro-level, by ascribing them to sheer spontaneity – i.e. not merely as *unpredictable* events – (unpredictable because of limits of quantum-interactions in measurement, or unpredictable because of insufficient information) – but as *uncaused* events, or absolutely spontaneous events. Thus, the quantum fluctuations, like Epicurus' *clinamen*, would violate causality, and be taken as 'free' or 'spontaneous' occurrences, in principle unexplainable.[7] This second alternative – apart from its vulgarization as the 'explanation' of free will in man – often is phrased as an argument for the 'creativity' of matter, and as an explanation for 'real novelty' in nature.

What is interesting about these two alternative 'escapes' from the consequences of a materialist monism is that they are metaphysically opposite to each other: the first argues for the absolute inertness of matter, and calls in a *Deus ex machina* to account for action, motion and change; the second makes matter self-active only in its spontaneous events, and presumably, all the laws of motion and change derive from this 'spontaneity' by statistical derivation, through time. (This was Epicurus' and Lucretius' view, and was also developed by Peirce, in his 'Tychistic' natural philosophy, in which he *derived* a growing determinism from an original pure chance.)

I would propose, instead, that the notion of 'First Cause' *and* 'chance' are, finally, limited anthropomorphisms, deriving from the social and historical experience of mankind *at a certain stage of history*, and therefore, from a *certain relation to the material universe*. Indeed, I would go further, and say that the notion of 'externally-acting forces' upon an inert matter is itself derived from the social-historical models of *human* action upon nature, taken as a passive receptacle, or upon other human beings, taken as passive instruments of another's will, i.e. as objects of exploitation. But that the notions of interaction among 'internally-related' constitutents of a system, which I argued for as a consequence of certain views in modern physics, goes *beyond* these other models, in seeing nature as inherently a self-active *system*, in which human agency is only one evolved mode of this self-activity.

It may be argued that even if this is true at the micro-level, it is false at the macro-level, where particulatesness and discreteness of entities and 'external relations' among such entities is not only a practical matter of our experience, but is even built into our very perceptual apparatus and our bodily structures. I think this is true. But it is not the particulateness

of absolute 'inner-outer' boundaries here any more than at the micro-level. Rather, it is the relativized particulateness of internally-related constituents of a system; or the particulateness of the internal structure of the material world.

<div align="center">III</div>

What epistemological and methodological consequences are there from such a view? Epistemologically, it would seem that knower and known, if they are internally related in the act of cognition, constitute a system, (or a subsystem of the larger system) and that the conditions of observation, experiment and thought depend as much on what is observed or experimented with as they do on the observer or experimenter. The *fact* of knowledge – and of scientific knowledge in particular, is therefore an *interaction*. I am not saying here that our knowledge is a knowledge *of* interactions; but rather that the very fact, the *material* fact of knowledge is itself an interaction. In short, no subjects without objects, in knowledge; and insofar as the object of knowledge, in this interaction, is a material constituent of a system, what the subject *knows* in such an interaction is the material world, insofar as he interacts with it, and in the mode of his interaction. But this is not to say, either (1) that the object exists only in this interaction or (2) that the object is passive in this interaction, and only the subject is active. Rather, as constituents of a material system, the subject *and* the object exist in the system as material entities; but are distinctly related to each other in the specific activity (or interaction) of knowing. Also, the object *in* this relation is no less self-active than *out* of it; but this self-activity now is mediated by the activity of the subject, in the interaction, just as the subject's self-activity is mediated by the activity of the object.

Thus, there is no claim that things exist *only* in relation knower-known, or observer-observed; but that as they interact in *this* system (or subsystem of the larger system) their interaction is a relation between one self-activity and another. 'Inertness', in its relative form, becomes a matter of perspective only.

The methodological consequences, for scientific inquiry, of such a view are *first*: that the absolute ontological breach between macro-world and micro-world is eliminated; rather, what is seen as anomalous is the imposition upon the micro-world of the concepts and prejudices of an historically-developed conception of the macro-world, deriving from a

particular mode of human praxis. Such a view does not yet give us the alternative concepts of the macro-world, nor does it solve the problems of theory-formation in micro-physics. But is has a certain heuristic force: namely, that the conception of interaction among internally-related constituents of a system, with richly differentiated internal structure, which we have explored here in its micro-world context, may be a useful model for the reconstruction of our macro-world conceptions. Second, as a methodological consequence, such a view argues *against* (or is, in Lakatos' terms, a *negative* heuristic, i.e. it forbids) a conception of physical inquiry as an interaction between mental and physical consituents of a dualistic system of mind and matter, but rather sees physical inquiry as itself a case of systematic interaction in the material world-system, and thereby delimits subjectivist and idealist interpretations of scientific knowledge. The constraints of scientific inquiry, thereby, become the constraints of the *conditions* of interaction, i.e. of observation, experiment, and theory; and it is the function of the theoretical imagination to invent and discover *new* conditions of interaction to extend our knowledge. Such conditions are not simply new experimental apparatus, in the sense of laboratory apparatus. Rather, the mind of the scientist is itself an 'experimental apparatus', using the 'equipment' of abstraction, imagination, mathematical construction, *Gedankenexperimente*, as its means of conceiving and creating such new conditions of interaction. And these very modes of conceptual activity, I take also to be the complex interactions of a (relatively isolated) subsystem of the material world – i.e. science.

In this sense, it is the kind of 'metaphysical' or 'ontological' conceptual experimentation which provides heuristic power to scientific thought and which performs, in analysis and construction, a significant role in scientific inquiry itself.

NOTES

[1] T. Kuhn, *The Structure of Scientific Revolutions*. Kuhn's own ambivalence mirrors that of Sir Karl: "We both (Kuhn and Popper) insist that scientists may properly aim to invent theories that *explain* observed phenomena and that do so in terms of *real* objects, whatever the latter phrase may mean." (in *Criticism and the Growth of Knowledge*, (ed. Lakatos and Musgrave), Cambridge University Press, 1970, p. 2). But of course everything *depends* on what the "phrase may mean"! So much for ambivalent realism. . . .

² J. Agassi, 'The Nature of Scientific Problems and Their Roots in Metaphysics', in *The Critical Approach to Science and Philosophy*, (ed. M. Bunge), Free Press, Glencoe, Illinois, 1964, pp. 189 ff.

³ M. Wartofsky, 'Metaphysics as Heuristic for Science', **III**, (ed. R. S. Cohen and M. Wartofsky), in *Boston Studies in the Philosophy of Science*, D. Reidel, Dordrecht, 1968 pp. 123–172. Reprinted in this volume, pp. 40–89.

⁴ P. Feyerabend, 'Consolations for the Specialist', in *Criticism and The Growth of Knowledge*, (ed. Lakatos and Musgrave), Cambridge University Press, 1970, p. 211.

⁵ See, for a discussion of this, John Stachel, 'The Rise and Fall of Geometrodynamics', presented October 1972, PSA meeting, Boston, in *Proceedings of the PSA, 1972, Boston Studies in the Philosophy of Science*, **XX**, (ed. by K. F. Schaffner and Robert S. Cohen), D. Reidel Publishing Co., Dordrecht, Holland, 1974, p. 31.

⁶ There may be sharp boundaries without field discontinuities, if non-analytic functions are admitted. See, for example, Bishop and Goldberg, *Tensor Analysis and Manifolds*, Macmillan, New York, p. 20.

⁷ In a stochastic account, of course, such 'fluctuations' are taken to be statistical phenomena, i.e. of 'mass-events' and *not* as 'spontaneous' unique events, for such unique events have no description in a stochastic account.

TOWARDS A CRITICAL MATERIALISM

[1971]

For more than a decade there has taken place an intensive discussion in Anglo-American philosophy on what may be characterized as a materialist approach to the Mind-Body problem. More specifically, the discussion has centered around the question, "Can sensations be reduced to brain-events or brain-processes?" or "Can psychological events be identical with physical events?"[1] This present paper is a critique of much of this discussion, not because it is materialist but because it represents a systematically inadequate materialism. My critique will proceed, first, from a characterization of the *Problematik* to which this recent materialism addresses itself; then, to a discussion of the historical and social contexts in which philosophical problems arise, and specifically, the contexts of the mind-body problem. This raises the question of how to address oneself to a philosophical problem, and in particular to the problem of a materialist resolution of the mind-body problem. Finally, I will counterpose what I call *critical* materialism to the contemporary forms of what I call *analytical* materialism, in a discussion of the question of the identity and the materiality of mental events.

The thesis of this paper is simple: the recent moves in the direction of a materialist formulation of mind-body identity are inadequate because they all accept the traditional mind-body dualism as their problem. Its solution is therefore bound to this very setting of the problem. It is this approach to the Identity-theory that I will call analytical materialism. Instead, I propose to reformulate the problem initially in a materialist monist way, namely: How can organized matter think, feel, dream, wish, etc.? In effect, then, I wish to shift the grounds of the present discussion, or to effect a problem-shift, as well as a shift in the methodology of the philosophical discussion.

Several questions arise, therefore:

(1) On what grounds (methodological or metaphilosophical) can one simply *abandon* the problem of mind-body dualism, as I propose to do?

(2) In what sense do the contemporary identity-theorists and analytical materialists (e.g. Putnam, Feyerabend, Smart, Sellars, Rorty, Cornman, *et al.*) accept the *Problematik* of contemporary or traditional mind-body

dualism as the context of their materialist 'solutions'? And why is this bad?

(3) What does the distinction I propose between 'analytical and 'critical' materialism come to, with respect to the mind-body problem?

Before addressing myself to these questions, I want to shift ground not only with respect to the substantive issue between dualism and monism, but perhaps even more radically, in order to establish a difference between analytical and critical materialism. I want to start with a problem-shift in a methodological sense, which reformulates what criteria for a philosophical analysis should be, and what such an analysis should include. This shift in itself will be part of the argument against analytical materialism.

It would be a normal expectation to have a philosopher examine critically the validity and substance of his opponent's *arguments* or his *premises*; to have him explicate and make clear the meanings and entailments of his own statements; to engage, therefore, in that noble practice, initiated by Socrates, of *dialectic* – i.e. of rational critique, according to the canons of non-contradiction, clarity, valid inference. Moreover, in constructing his own arguments, one would normally expect a competent philosopher to be coherent and systematic, and one should hope that what he says will be interesting.

I don't mean to eschew these philosophical virtues. They are the marks of our craft and the conditions of competence. They constitute, in effect, the grammar and rhetoric of philosophy. Overriding this competence, however, is the question of truth. Just as one may speak grammatically, and even persuasively, and yet utter false statements or trivial ones, so too one can be a competent philosopher and a persuasive practitioner of argument and yet end up with nonsense or falsehood or triviality. The strongest moral imperative in philosophy is to strive for truth by way of inquiry. And it is notorious that the very problem of truth, its definition, its criteria, its conditions and its modes of utterance, is itself one of the thorniest in philosophy, and constitutes, in effect, that issue in which philosophy is self-critical of its own foundations and its modes of practice. Thus, this very problem of truth, both in its formal theory in logic and in its substantive modes in the theory of knowledge, of belief, of practice or action, lies at the very heart of philosophy.

This digression brings me to the point, to be made briefly here: that in philosophical inquiry – specifically, in that mode of inquiry in which one assesses alternative theories critically – a crucial aspect is the examin-

ation of the social and historical contexts of the theories one is assessing.
This is based on the grounds that philosophy is practiced by men, and
that human activity of any sort, including philosophical activity, is both
product and agent of human society and human history. The premise
here, then, is that genetic analysis bears on philosophical issues, and that
the nature of a philosophical problem is in part contextual; that what is
at issue, and what will count as a solution to a given problem is deeply
rooted in the historical, social and scientific contexts in which the
problem is ·raised and understood.

What follows from this is the beginning of an answer to at least two,
and perhaps to all three of the questions I initially posed: (1) On what
grounds can one abandon the dualist framework? (2) In what sense do
contemporary Identity-theorists accept this framework? (3) What does
the distinction between 'analytical' and 'critical' materialism come to? I
will argue that mind-body dualism is itself an historically-bound theory,
developed in a limited set of historical (hence also scientific, cultural and
social) frameworks which are now anomalous; and therefore, that its
persistence in analytical and philosophical contexts is also an anomaly,
fostered by the relative autonomy of philosophical thought – i.e. by the
tendency of philosophers to deal with the history of philosophy or of
philosophical ideals *a*historically, as simply inherited problems or as
contemporary problems; and therefore to abstract the philosophical
contexts from those real contexts beyond philosophy in which such
problems originally had their roots, and to which they are relevant at
present. The gap between the original and the contemporary contexts,
once ignored, leaves the problem open to formal or abstract analysis.
Therefore, an analytical materialism, which argues for Identity-theory,
even if it uses the most contemporary formal apparatus of analysis, does
so as if the dualism which it criticizes is a timeless analytical problem to
be dissolved or solved by analysis as no more than. a problem in the logic
of concepts or in the analysis of meanings.

The consequence is that the historical sources of this dualism are
overlooked, and the dualism itself is taken in the narrowest of its
analytical formulations, as an abstract problem: namely, in its
ontological mode as the problem of the relation between an inert matter
and an active, agential thinking self.[2] Or more narrowly yet, as a
question of linguistic or conceptual reduction, bearing on issues of
translatability and the eliminability or non-eliminability of mental or
psychological terms. The analytical materialist critique, therefore, takes

on the characteristics of what it criticizes (as the ancient conquerors of China were said to become 'Sinified' in short order) and remains a narrow and abstract mechanistically conceived materialism in its ontological modes, and an emasculated theory of language and of conceptual analysis in its linguistic modes.

Beyond the competence of philosophical analysis, therefore, there lies the problem of an appropriate problem-setting (*Problemstellung, Problematik*). Let me try then to dissolve the older framework first and then present a new one.

The old or traditional mind-body distinction rests on the view that whereas matter, or body, or corporeality is inert, mind is active, and in effect self-active. Certainly, on the traditional view (e.g. of 17th and 18th century mechanical physics and its natural philosophy) matter is in motion, but this motion is not its own. It is imparted motion, and therefore has its source outside matter – ultimately, by the logic of this argument, in a Prime Mover. The newer mind-body distinction, on the other hand, concerns the irreducibility of mental activity or of mental or psychological events to physical or physiological activity, or to 'brain events'. An outright dualism argues that, no matter how far our knowledge of physical or brain function proceeds, it cannot explain or give an adequate account of such mental acts as thinking, feeling, willing, deciding, intending, etc. – in short, of conscious, intentional actions on the part of an agent. The reason this cannot be achieved, it is argued, is that no physical account can include the element of agency or conscious intention required for the adequate characterization of mental acts, or the element of awareness required to characterize even sensations; and this is because physics quite properly does not include either such agency or intentionality or awareness as a feature of physical description, or of physical theory. This outright dualism has traditionally led to either a clearcut mentalism or idealism; or, as the other side of the coin, to a reductionist materialism in which 'central processes' or mental life was simply denied or eliminatively reduced (as in crude behaviorism).

The analytical materialist critique of this mentalism takes off from a somewhat modified and more modest version of this dualism, which holds that there may indeed be a physical event (or complex of events or processes) either contingently or necessarily associated with every mental act, as its basis or as its condition; but at most, there is only a correlation or a parallelism of mental and physical events. The two types of event remain irreducibly distinct, the mental events requiring a description in

terms of agency or awareness which a physicalist account cannot give. The aim of reductionist materialist Identity-theory is to get beyond this psycho-physical parellelism, i.e. to propose that the 'two' events are really only *one* event, under two descriptions – that there is an identity of the mental and the physical in fact, though one may view this identical event under two aspects, or describe it in two alternative languages, one physical, the other, mental. (Thus, it is a version, though a simplified one, of Spinoza's 'double-aspect' theory of Identity). The *strong* version of Identity theory holds simply that the 'mental' language is mistaken altogether, and that it is a holdover from more primitive animistic thought, from 'soul'-theology; and that the identity of what is called 'physical' and 'mental' therefore bespeaks *not* an identity of two things, or events, or even two alternative descriptions of the same thing, but rather an 'identity' (or self-identity) in terms of brain-events (as physical or neurophysiological events) under only *one* proper description – namely the physical, or 'scientific' one. The 'mentalist' description is therefore to be eliminated as an essentially unscientific description, and terms of reference in that description, such as 'pain', 'wish', 'thought', 'intention' are to be understood or explained as complex brain-events. Thus the *facts* of so-called 'mental life' are not eliminated. People still continue to 'have' what we call 'pains', 'intentions', 'thoughts', etc., but the proper scientific reference of these terms is to a brain event or a complex of such events.

   That's as far as analytical materialism goes, though the arguments are more complex and often of great subtlety. In short, it accepts the problem set by mind-body dualism, and overcomes it in terms of mind-body identity, on the grounds of the identity of physical events or 'brain-events' and mental events. (There are a few exceptions which go beyond this, in the direction I will also urge, but I will mention these later).

   Now let me shift the problem, by dissolving this framework. Philosophers and scientists didn't always hold that there were two realms, the mental and the physical, which somehow had to be brought together or correlated. Dualism is indeed an early view, deriving, in part at least, from the historical separation of ideational and mental function from physical or bodily function. This separation took place in fact, in social life, before it took place in the minds of philosophers.[3] In ancient society, the work of the hand, of the body was very early socially separated from the work of the head, or of the mind. Discourse,

including the storing of legends, of myth, of history, as well as the formulation of rules and laws, became the function of a ruling class, whether of priests or kings or legislators. The practical work of production, of building and maintaining the physical basis of society, became the work of a ruled class – whether of artisans, or farmers or slaves. This separation of head and hand, crudely sketched here, is the source of the philosophical reflection of this separation in theory. There are other so-called 'natural' sources of the mind-body dualism, which build upon this separation. For example, the interpretation of dream life as constituting a non- or super-natural realm; the myths of life after death, with their requirements for a persisting, though disincorporated or disembodied consciousness. But it is interesting that even these instances have been given materialist interpretations, in the early mythical materialism which conceived of dream bodies as real physical entities, and of life after death as a corporeal or bodily existence. (The 18th century scientist Joseph Priestley retained a 'materialist' theory of immortality of the soul: since, he argued, there is no consciousness without the material body, and since there is conscious life after death, then there must be a bodily resurrection after death![4]) The dualist interpretations arise either out of the head-hand split; or out of the distinction between appearance and reality, between real and unreal worlds, a distinction which comes early in Greek philosophy. In this latter epistemological context, however, one of the two poles of the dichotomy is seen as real, and the other as unreal – therefore, *only* a false appearance, or in effect (since all appearances are taken as 'false' on this view), as a mistake in judgement. Thus, for Platonic idealism, as for Berkeley, the ideational world is real, the physical world only a shadow-world, a 'moving image' or copy of this real world. For a Democritean materialism, as for Hobbes, only the physical world is real, in its moving atoms, the mental world a phantasm, an *eikon* or an image of this material reality.

Suppose one were to accept this (oversimplified) account of the genesis of dualism. Why should one then reject dualism? Isn't such an argument for the rejection of dualism simply a transparent form of the 'Genetic Fallacy'? Let me approach the question from a different angle. Suppose, now, that we take a radically alternative premise to dualism, i.e. an outright materialist monism, and ask how such a theory would pose for itself the problem or the question of the nature of thought or mental activity. The advantage of such a view was already clearly seen by

Aristotle, who argued (in the *De Anima*) that once the mental and the physical were separated, no theory of their interrelation would be adequate; and that therefore, the separation should not be made in the first place. Now, philosophically, this is not a genetic argument, but only a conceptual or analytical alternative. It is the countertheory. One may conceive of it simply on rational or conceptual grounds as an alternative conception of the problem, one which goes directly contrary to Platonic dualism. In the latter, thought, or the Ideal Forms are independently real entities, and body and matter are simply the images of these forms, (or in the more complex and problematic account in the *Timaeus*, the potential receptacles, which receive the forms). Put in this way, these two views are simply conceptual alternatives. But the question for a critical philosophy, – i.e. a philosophy which is historically as well as analytically self-conscious – is: How did such two diverse frameworks arise within the same contexts of Greek life, Greek science and culture? Thus, the genesis of a philosophical problem is seen as relevant to its understanding, if not also to its solution; and the question of *Problemstellung* becomes a relevant philosophical question.

As no more than a programmatic sketch for such an approach, one might propose that the differences between Platonic and Aristotelian frameworks, (acknowledging their deep similarities) derive from differences *within* the same historical and cultural setting: differences such as that between Plato and Aristotle not only as philosophers but as historical human beings in a complex transitional society. So, for example, the Platonic tradition is tied to that of the Athenian aristocracy, apprised traumatically of its political role by the defeat of Athens in the Peloponnesian Wars, and by the desperate political flux of its aftermath. Thus, Platonism adopts the discipline of thought, or reason as its ideological substitute for a defunct mythology, and develops the tradition of a changeless mathematical reason as its paradigm of stability, order and justice. The Aristotelian tradition is that of the craftsman-scientist, wherein reason is the essence or the norm of *praxis*, of human action in the making of things, of *poiesis*, and wherein *psyche* is the form of the normal activity of natural living things in their development. The use of the concept 'Form' is therefore different in Plato and Aristotle, and much of Aristotle's critique of Plato is over this concept, just as, at the same time, Aristotle's framework depends on the dialectic with the Platonists. Their different views of the relation of form to matter, of mind to body, are not simply conceptual alternatives in

some context-free domain of 'philosophy', but rather the rich and extraordinarily complex context-bound reflections of a concrete history and culture, filtered through the genius of the individuals, and through the very developments of linguistic and logical forms introduced by the philosophical schools and by Greek literature and poetry. The 'Platonism' and the 'Aristotelianism' inherited through the history of philosophy is itself a historically complex transformation of the concepts and language of the originals, and the conceptual issues between these schools, as *we* receive them, bear the imprint of both the historical and contemporary contexts. *Which* mind-body problem are we then addressing? How shall *we* formulate the problem: as a problem within the relatively autonomous dialectical tradition of the historical formulations? Which ones? The Greek, or the medieval Arabic and Jewish transformations, or the theological forms which Latin Europe imposed on the translations from Greek and Arabic? Or have we formulated our own problem, free of historical complications, in the 'neutral' scientific or logical language of 'philosophical analysis'? Or is the 'problem' simply one concerning how 'we' use certain terms and certain linguistic forms?

The programmatic sketch I gave above was crude, and obviously needs refinement, criticism, analysis. So too, the questions I have just asked are both too broad and too narrow. I offer them simply by way of a suggestion as to what does need to be taken into account in establishing a setting of the philosophical problem. What I propose, in short, is that a historical sociology of the history of philosophy is a relevant way of getting at the question of *Problemstellung*. Specifically, the mind-body problem is on the one hand seen in its context as a problem *within* the history of philosophy; and on the other, it is seen as a contemporary problem, *in this light*. A critical approach would assess the alternative models of mind-body relation in this setting. For example, the scientific model of much of the contemporary discussion of this problem *remains* that of 17th century mechanics and physiology, as it derives from Cartesian and Lockeian frameworks. But 20th century physics and neurophysiology operate with such anomalous models only at a simple engineering level, and have long abandoned the models of classical mechanics, optics, physiology. Common sense does operate with such models, because common sense is essentially that which has been formed by classical 17th century physics, in its molar views of matter and action. Theoretically, such models have long been anomalous. Yet much

philosophical discussion, especially in epistemology, e.g. on perception, and on the mind-body problem as well, proceed as if these anomalous models were still viable. Another anomaly or conceptual incoherence arises from the combination of powerful new formal modes of analysis, e.g. in advanced 20th century mathematics and logic, with primitive mechanistic engineering models of physical process, or even more primitive models of biological process.

The fault in analytic materialism – or at least, in much of it – is that it abides these anomalies unconsciously – i.e. uncritically – and thus generates both historically as well as conceptually anomalous problems. As an example of this, let me take the most characteristic problem-formulation of the Identity theory in analytical materialism: 'Can mental (or psychological) events be reduced to brain-events'? or 'Are sensations brain processes?' The anomaly is in thinking of conscious or mental activity of any sort as a 'brain-event' or a 'neurophysiological event' *tout court*. In short, analytical materialism is bound to a model, or to models of thinking, or of mind, which are *categorically* wrong. Whichever solution or proposal on Identity may be forthcoming, *two* things will be wrong: (1) The 'Identity problem' itself stems from a problem-setting which *begins* with the framework of mind-body duality, especially in its mechanistic 17th century forms, and so its solutions are solutions to the wrong problem. (2) The identity will be an identity at the wrong level, categorially, for the pair 'mental-event'-'brain-event' are an anomalous pair categorially, like 'cat' and 'justice'. This is because the concept 'mental' takes as its domain a concept of 'physical' or 'material' irreducible to 'brain event' or 'neurophysiological event', on any conceivably viable Identity theory.

But this seems for all the world as if it were a rejection of materialism. If 'mental event' is not to be taken as identical to 'brain event', where the brain is taken as the material organ of consciousness, what is left for materialism? The burden on a critical materialism would be two-fold then: first, to show that the *Problematik* is wrongly conceived or is anomalous, because *a*historical and conceptually formalistic (as I have suggested above); but second, to show that this 'brain-reductionism' is iteself an inadequate materialism, and to suggest what a more adequate materialism would be.

The question then is: If I give up the basic material particulars which a physicalist or neurophysiological materialism rests upon, am I not giving up materialism as well? If 'brain event' or 'neural event' are *not* to be

identified as the materialist's options for reductions of, or replacements for 'mental event', what is left for materialism? This is precisely where a critical materialism and the typical analytic (reductive, eliminative, or even 'adverbial') materialism part ways. Let us start with the case of 'sensing': it is certainly the case that in tactile, visual, aural, olfactory, gustatory, kinesthetic or proprioceptive 'sensings', brain events are involved. But so too is the whole network of the sensory process, from afferent and efferent neurons, through the mediation not simply of some undifferentiated lump of tissue called the 'brain' but also the highly complex *system* including the problematic reticular formation, the hypothalamus, the various centers of the cortex, etc. etc. Very well, then, let's expand the notion of 'brain-event' to include the whole systemic event, as a neurophysiological event involving much more than the 'brain' by itself. But beyond this, as a systemic response, the very nature of the sensory event involves the whole bodily disposition, the condition and 'set' of the physico-chemical properties of tissue, muscle, fluids, blood, enzymes, etc. These are coordinated, in the normal sensory event, in a whole organism,[5] say in a person's body as a whole, or in an animal's body as a whole. But beyond this – and this is crucially missing from analytical materialism's Identity theories – the mental event – even the simplest sensory event – is bound up with a *biography* and a *history*. By the first, I mean *ontogeny* – the course of development of an individual, intact and continuous organism; by the second, I mean *phylogeny*, i.e. the species-history of that organism, in both its biological and (in the case of human beings) its cultural and social evolution. We may in fact abstract the situation for purposes of inquiry. Thus, for example, we may study sensory phenomena in decorticate organisms, in order to abstract and separate the parts of the total process, just as we may also study single neuron firings in sensory processes, in order to analyze the constituent anatomical and physiological elements of the sensory process. But what distraction it would take for a scientist or a philosopher to call *that* a case of 'sensing' or a 'sensation'! A sensation is not something that happens in a neuron; nor does it happen in a brain. It happens in an organism, whose intactness and systemic integrity is a necessary condition for the very occurrence of the phenomenon of sensing. That we may abstract and analyze the components of the process is clear enough. Confusion begins when part is taken for whole, and abstract particular is taken for concrete individual. So far, then – whether for people or other animals (to borrow from the title of a recent

article by Marjorie Grene[6]) sensing is a phenomenon of integrated organisms with nervous systems. But this is far from the whole story. In fact, I am just warming up. Beyond this, sensations belong to such organisms, i.e. they are subject-bound, and cannot exist otherwise. (Charles Dickens provides the best *reductio ad absurdum* here. In his novel, *Hard Times*, Mrs. Grundy is asked by her daughter, "Are you in pain, dear Mother?" to which the doughty old lady replies: "I think there is a pain somewhere in the room, but I couldn't positively say that I have got it.'") The point is, of course, that it is individuals or persons who sense. Brains don't. Sensing (and certainly also thinking, perceiving, imagining, desiring, etc.) is *mine* or *yours*. For any species-individual, its conscious activities are already not simply its own, however, but are also mediated phylogenetically. The frog's eye (as Lettvin and Maturana have shown), is a *frog's* eye – it is bound in its very structure and function to a frog's world, and is evolved and differentiated phylogenetically. Insofar as you and I share the natural world with the frog, there is some intersection of frog and human vision, for example, but only *some*. Thus, the very organs of conscious activity are forms of life, i.e. forms of practical interaction of world and subject, which are evolved, adaptive and selective. Marx's striking phrase about the eyes and ears, as "organs that tear man away from his individuality, converting him into a mirror and echo of the universe"[8] (and similar expressions, e.g. in Feuerbach, in Leibniz) suggest even phylogenesis as itself not a self-determining system, but as the product of the evolution of matter itself, incorporating into the species the historical or evolutionary forms of its whole development. But therefore the very condition of the neurophysiological process itself – the so-called material basis of consciousness – is set into a larger context than either 'brain' by itself, or 'organism' by itself. Even the lowest act of sensing is already a *species*-act, and its 'material basis' can be nothing less than the species, in all its complexity of evolution, interaction, and, in the human case, cultural and social development in history. Then why all the silly reductionist business about 'mental events' as 'the firing of C-fibres', which echoes through the muddled discussion of the Identity theory?

Consciousness, even at the biological or physiological level, is already both ontogenetically biographical, and phylogenetically species-and-history based. The very content and fact of a sensation is thus mediated by an individual biography, with its accretions of selection, attention, responsiveness already a part of every act of sensing, providing the

background for the specific situation. How much more clearly is this also the case in higher mental function! Thus too, every act of sensing, not to speak of imagining or thinking, is bound up with a species-history, determing the range and specificity of the object domain, the character of the organ system and its parts, down to the neurons, tissues and blood cells.

What are we to say of the human case, however? After all, noone has gotten terribly excited over the mind-body problem for frogs and fish, or even for the apes. One non-reductionist approach is to talk of *persons*. This hides as much as it reveals. For the *Person* may become as confusing an abstraction, as *a*historical and reified a concept as 'brain-event' or 'mental-event', and as ontologically opaque an abstraction as 'matter' or 'mind'. Yet it has the initial advantage of characterizing a complex and distinctive organism, namely, a self-conscious one, possessing language. Permit me a few additions or ramifications of these general properties: socially-organized, culturally transmissive, educable – in short, a historical being of *praxis*, of social human action. Immediately, in this context, the 'brain-event' becomes proliferated beyond the physiological, and both ontogeny and phylogeny become social, cultural, historical, linguistic, far beyond the limits of the physiology of cell-tissue. In short, the sensation *I* have, the thought *I* think, the desire *I* express, the action *I* perform as a human being is hopelessly infected with my personal biography, my species-history, my social and historical past and present *and* future. For simple sensation, I may be bound only to my biological ancestry and my own growth patterns, and even these are, I fear, not enough to fully characterize a human sensation, since I take such a sensation to be social in a way no less than, but quite different from that, say, of human speech, or courting behavior. But that is a very problematic and ultimately a scientific-technical question for further inquiry. But if we are indeed talking about such higher functions as thought, for example, it is almost incomprehensible to me how the distinctive human parameters of such complex conscious events or processes can be conceived apart from this ramified context.

Here, the reductive or the eliminative materialist has an option: He can claim that *all* of this phylogenetic-historical matrix is patterned *in* the brain-event; that 'physiology recapitulates sociology', if you like, and that the history of the species exists *materially* in the tissue itself. Such a reduction is tempting, because it permits one to have one's socio-biological cake and eat it too. But for too much of our

species-existence, our organs are simply transducers and mediators, not storage-cells, though the very structures of these organs do embody the phylogenetic history of our species adaptation at the biological level. Beyond this, we embody our social-historical existence as a species in language and social artifact, coded so that we can communicate with it by means of our organs of consciousness, i.e. by our senses, our perception, feeling and thought. Much of our 'software' has become 'hardware' in the economies of technology and culture, for purposes of storage and transmission. The brain is, by this means, 'enlarged': or as a physiological organ, relieved of these storage and transmission functions.

But what about the brain? Are materialists hereafter to give up the brain as the material organ of consciousness, in favor of some diffuse organicist notion of 'matter' spread through time and space across species and cultures and histories? Doesn't this abandon a 'hard' for a 'soft' materialism? I think not. The brain is, after all, the *organ* i.e. the *instrument* of consciousness, and of thought. But precisely because it is the *organ* of thought, neither it itself, nor 'events' in it are identical with consciousness or thinking, no more than 'foot-events' are 'identical' with walking. My feet don't walk. *I* walk, using my feet. 'I' could no more walk 'without' my feet, than my feet could 'walk' without me. But this grammatical nonsense simply reveals a fallacy of analysis, a categorial confusion. The 'I' here is not some Machian bundle of sensations, nor yet a Humean 'common center' of sensations, nor yet a 'transcendental Ego', but rather the concrete event of 'my body walking'. But 'my body' is a human, i.e. a species-body, though an individual, a concrete 'I'. And walking is a species-act, which 'I' do, individually. Thus too, my brain doesn't think; rather, *I* do, by means of my brain, as the most specialized organ in a system of organs constituting my body, i.e. constituting me as the concrete individual *identified* with this body, its history and its activity. One could say, then, my *body* thinks – but only as a human body; and therefore the mental act, the act of thought can only be carried out by a human body. The act is thereby a human act, i.e. an individuated species-act. The identity is no longer a 'mind-body' identity between two things, a mind and a body; rather it is the self-identity of one complex material organism, i.e. it is 'thinking-matter' but in the only way that we know of, i.e. as a highly organized and evolved individual human being.

A physicalist reductionism cannot give this categorial analysis; but not

because it is physicalist. Rather, because it is reductionist. Neither can a materialism limited to either the framework or the *Problematik* of the traditional mind-body problem in its dualist formulation. And certainly, an analytical materialism for which the problem-setting remains in terms of 'flip-flops', 'C-fibres' and reductively conceived 'brain-events'; or for which the issue remains *only* at the level of 'the logic of identity statements', cannot give the account adequate to the complexity of the phenomenon itself.

A critical materialism eschews both the dualist *Problematik* and the analytic reductionism of either the physicalist or linguistic varieties; and the latter, not because either conceptual or linguistic analysis *as such* commits one to an ahistorical and formalized abstraction of the problem, but because the conditions of the methodology have been mistaken for the conditions of the problem itself. No evolution, no history, no society – no mind. Therefore, without these parameters as *materialist* parameters, analytic Identity theories become irrelevant to the formulation of an adequate materialist theory of mind, and may indeed become a philosophical *cul-de-sac*.

NOTES

[1] Many of the important essays in this discussion are collected in *Materialism and the Mind-Body Problem*, David M. Rosenthal, (ed.), (Prentice Hall, Englewood Cliffs, N.J., 1971), which also contains a thorough bibliography of the discussion.

[2] I have discussed this in several other contexts. See, e.g., my 'Diderot and the Development of Materialist Monism', *Diderot Studies II*, Fellows and Torrey, (eds.), (Syracuse University Press, 1953), pp. 279–329; 'Spinoza on the Passions – Towards a Scientific Psychology', *Spinoza*, M. Grene, ed., (Anchor Press/Doubleday, 1973), pp. 329–353; and more recently, in 'Matter, Action and Interaction', *Proceedings of the XV International Congress of Philosophy, Varna, 1973*.

[3] Both Marxist and non-Marxist scholars have argued for this thesis. See, esp., B. Farrington, *The Separation of Head and Hand in Ancient Greece*, as well as *Greek Science*; and also, John Dewey, *Reconstruction in Philosophy*. For a full discussion of the 'natural' sources of dualism, in psychological and epistemological terms, see A. O. Lovejoy, *The Revolt Against Dualism*.

[4] J. Priestley, *The Doctrine of Bodily Resurrection Upheld*.

[5] This point, of taking the whole organism as the appropriate material correlate for the 'mental event', is made within analytical materialist discussion, e.g., by T. Nagel, 'Physicalism', in D. M. Rosenthal, *op. cit.*, pp. 98–100; and by Hilary Putnam, 'The Nature of Mental States', *ibid.*, p. 154. These are welcome exceptions to the usual view, and show that analytical materialism is not a closed system, and can become critical. But the grounds for these exceptions are not yet critical grounds.

[6] M. Grene, 'People and Other Animals', *Philosophical Forum*, **III**, 2 pp. 157–172, (1972).

[7] Charles Dickens, *Hard Times*, (The New American Library, New York, 1963), p. 198.

[8] K. Marx, F. Engels, *Sochineniia*, Vol. I. (Moscow, 1929), p. 180. (Cited by Loren Graham, *Science and Philosophy in the Soviet Union*, New York: Knopf, 1972, p. 386).

THE RELATION BETWEEN PHILOSOPHY OF
SCIENCE AND HISTORY OF SCIENCE*
[1977]

In the heat of a debate, it is often hard to see how one-sided and some-times foolish the opposed positions are. The debate on the relation of the history of science to the philosophy of science is so patently an un-reasonable one that one has to step back to see it in its full absurdity. Further, in order to understand why reasonable and often intellectually superior people have been caught in this absurdity, it will help to examine the alternative positions and to suggest how and why they developed as they did, both conceptually and historically.

The problem posed by the debate is the following one: *Two disciplines – philosophy of science and history of science – are both concerned with what is ostensibly the same thing – namely, science. Are the two related?* The problem in this form is vacuous. It is resolved vacuously in the way the question is put. If the two disciplines are *about* the same thing, then that is already *a* relation. But obviously a vacuous one. The issue is: *How* are they non-vacuously related? Or: *Are* they in fact about the same thing?

In this paper, I will argue:

First: that the relations are crucial ones, for both disciplines; crucial in the sense that a non-philosophical history of science is a fatally flawed history of science; and that an ahistorical philosophy of science is, similarly, totally inadequate to its task.

Second: that *such* inadequate and flawed history and philosophy of science do exist, and that the debate arises out of just these crucial inadequacies; but that its present terms do not lead to a resolution of the problem, but only confound it, or mystify it.

Third: that the conception of science, and of the activity of science – if you like, the ontology of science, what science is – is what is at issue; and that this conception needs to be recast. In brief, I will argue that science is neither simply a body of knowledge, in some stable form, whose anatomy can be understood by dissection and reconstruction. I therefore reject the rational reconstruction view as it is represented in

the dominant philosophy of science of the last half-century. Nor is it simply a sequence of practices, discoveries, beliefs or forms of institutionalization. Nor is history of science a chronicle of *facts* concerning all of these. But further: neither is science, or its history, simply an internal or rational dialectic of problems and solutions, conjectures and refutations, proposals and replacements. I therefore reject a series of views on the history of science, as either factual (external) chronicle, or internal dialectic, though I accept some of the fundamental premises of such historical views – namely, that science changes, grows, or develops; and that theoretical criticism and replacement of theories are crucial to this growth.

The alternative ontology of science which I will propose is that it is an historically evolved mode of human cognitive action or *praxis*, – and that as such, it is essentially a teleological activity. Its character or nature cannot be understood apart from its purposes *in this* context – i.e. its purposes or cognitive praxis in socio-historical context. Further, that such a teleological mode of action is essentially theoretical and that is what distinguishes it from other modes of action, including other modes of cognitive activity; and that the growth and change of scientific theory has both an internal theoretical dimension, and an external one, having to do with the genesis of theory, the uses or practices to which the theory is put, and the conditions under which theories change or are replaced. Such an ontology of science, therefore, construes science essentially as a mode of cognitive human action or praxis; as such, teleological; and distinctively theoretical.

Fourth: I will argue that such a view of science requires, as *its* adequate theory of science, a philosophical history of science, and an historical philosophy of science, though what these terms mean will have to be elaborated more fully later. Insofar as I will characterize science as *cognitive* human action or praxis, the emphasis is on science as a knowledge-acquiring activity, or more specifically as a truth-seeking activity. Philosophically, this establishes *epistemology* as the broader context for a characterization of this activity. But because this truth-seeking activity is to be understood in the context of its historical development, what is required is an *historical epistemology*. My own argument is therefore part of a broader program of what I will call *historical epistemology*, and which is to be distinguished from some of its look-alikes, now

emerging in philosophy and history of science – namely 'naturalistic epistemology' (Dewey, Quine, Campbell) or 'evolutionary epistemology' (J. Piaget, J. Herrick, Toulmin, Shimony, H. Yilmaz and to some extent Popper and Feyerabend). Broadly speaking, but in senses as yet unspecified, historical epistemology is Marxist in its orientation, and needs to be further characterized, especially with respect to other contemporary views which seem similar to it – e.g. those of G. Bachelard, Althusser, Foucault, and D. Lecourt, who represent some of the present discussion on these issues in France. But this remains to be done.

So much for the program of my paper and its context.

I

In order to examine the present debate on the relations of the history and the philosophy of science, it might be useful to do a typology of the relations which could be conceived to exist between them. It will be easily seen what historical and contemporary views fit these types, and I will sketch some of these examples, and their history. These views will be seen to fall into five types, the last of which I offer as an appropriate resolution of the problem. These are (1) AER, (2) DER, (3) AIR$_1$, (4) DIR, (5) AIR$_2$, (to be spelled out in due course).

(1) *Agreeable External Relations:* Since philosophy and history of science are *about* the same thing – science – the division between them is merely a division of labor. Philosophy deals with the philosophical aspects of science – i.e. the logical, epistemological and metaphysical aspects, and perhaps even the value-theoretical aspects – and history deals with the historical aspects – 'who did what, when, and with whom'. On this view, history and philosophy of science are *externally* related to each other, in an *agreeable* division. That is, each recognizes the other's domain, and accepts the territorial division, either explicity or tacitly. There is, however, no relation between history and philosophy of science of an internal sort. Not only do they not impinge on each other at all, but rather the concepts and modes of discourse of one do not effect the concepts or modes of discourse of the other. And of course there is no debate. This view represents a philosophical fiction, like the *state of nature* in political theory. It never existed but it is useful as a starting point, and makes the schema symmetrical.

(2) *Disagreeable External Relations:* Here, hegemony is claimed either
by one or the other or both of the disciplines, to the exclusion of the
other. In effect, the claim is that the other is fundamentally inadequate
to the task, or construes the common subject – science – in a wrong way.
What this means is that there are two competing *theories* of science,
mutually incompatible; and therefore, the debate becomes open. It is
not over the methodological relation of the history to the philosophy of
science, but whether science itself is to be construed in one or another
way. Yet the *outward* form of this debate may become one between
alternative *philosophies* of science, under the impact of historical con-
siderations; or between two theories of the *history* of science, under the
impact of philosophical considerations – and that in fact is what has
happened.

Stage one of the debate concerned the *rejection* of historical contexts
(as well as sociological and psychological ones) by the logical-positivist,
and logical empiricist schools of philosophy of science. It may be
necessary to characterize these views more fully later on, but for present
purposes it is enough to say that this school – let's call it *rational recon-
struction with an empiricist basis* – founded its view of science on the
division between *contexts of discovery* and *contexts of justification*, taking
the latter as its proper domain. Though this sounds like *agreeable* external
relations and division of labor, (assigning the contexts of discovery to
the history, sociology and psychology of science) it was not, and I will
say why later. Justification demanded (a) the rational reconstruction of
the *achieved* body of knowledge of contemporary science, i.e. the re-
casting of this body of knowledge – construed as a set of statements – in
the logical form of premises and conclusions in a system of valid
inference (ideally in axiomatic form); (b) an empirical or observational
base for the verification, or confirmation, or corroboration or accept-
ability of scientific statements. (The details are unimportant, and so are
the alternative verificationist and falsificationist alternatives here.)
Justification was therefore to derive from a universal and subject-neutral
way of linking theory to observation and experiment; and because
universal, it was conceived as *ahistorical.* However, there was tolerance
for alternative theoretical frameworks, as long as each could be con-
structed rationally and consistently, and could be linked by coordinating
or interpretive definitions to the empirical base. And there was also

tolerance as to how the empirical base could be chosen – i.e. as to what should count as protocol-sentences or basic predicates. This tolerance, however, led to a relativization of the reduction-base which has made the original project of justification bankrupt, at least on the terms originally conceived. (Details will be given upon request.)

The alternative *philosophy of science* which developed on the grounds of this very liberalization of the observational base itself, had its sources in two different contexts: (1) in the older philosophical relativism, rooted in classical skepticism, for which the simple notion of a subject- or context-neutral 'observation-statement' was always suspect. This view construed all observation as a function of context: e.g. interests and purposes, or conceptual frameworks, or cultural set. Being ahistorical, at least philosophically, this alternative conceptual framework had to rediscover this observational relativism in its available contemporary form, in the linguistic mode – namely in the Wittgensteinian distinction between 'seeing' and 'seeing as', or in its juicier versions, in Wittgenstein's "forms of language are forms of life" dictum – an incipient but under-developed sociolinguistic thesis already popular in the works of B. L. Whorf and Edward Sapir, and old hat to an old metaphysician like Collingwood. (I won't go into the venerability of this tradition here, since it is embarrassing.) (2) On the other hand, this alternative philoso-phy of science drew its sustenance from historical considerations: The philosophers of science had *discovered* (!) that theories change, that science replaces old theories with new ones, and that there is at least a problem in understanding the *growth* of science. The older notion that science progressed by incremental and cumulative improvements within the one true theory began to crumble before the historian's account of radical changes and replacements in theory.

This historical influence was aided and abetted by a naturalistic epistemology which claimed that, as a cognitive practice, theoretical scientific activity grew out of and was involved in the practical, social and psychological contexts of success and failure; in the animal behavior that scientific inquiry was rooted in and which had its basis in trial and error; and that science was also rooted in, and influenced by the social contexts of ideology, belief, and the common forms of group and institu-tional behavior. Therefore what entered into the fabric of this theoretical mode of social cognition which is science, were purposes and interests not

always conscious, not always rational and not always explicit. Such a naturalistic epistemology thus linked the analysis of scientific inquiry to the more general analysis of inquiry as a social and natural mode of human activity; and though such a naturalism was not yet historical; it took the arena of social and cultural history into account.

The further weakening of both the logicist and the empiricist dogmas – by internal criticism (e.g. by Quine and Hempel), by the discussion of the older analytic-synthetic distinction – all contributed to the growth of an alternative philosophy of science, (whether as Hansonian, Polanyian, Kuhnian, Feyerabendian, Lakatosian, Agassian, or Quinean. I will not go into the independent bases for this alternative view here, since they are treated elsewhere in this volume and in the current literature at great, if not enormous, length.)

Thus, within the philosophy of science, *Disagreeable External Relations*, in stage one, generates a split between two philosophies of science – one of which (rational reconstruction) had become all but canonical, in its ahistorical mode; the other of which threatened to usurp this hegemony, with a peculiar link to history of science.

Stage two of D.E.R. concerns the history of science itself: its disagreeable external relations were with the logical-positivist, logical-empiricist philosophy of science, which it took to be as irrelevant to science (and not just the history of science) as the philosophers took history of science to be irrelevant to justificationist philosophy of science. Now to be fair, rational reconstruction was not opposed to history of science, and regarded it as useful to know – if not as a history of past mistakes, then at least as a resource for examples, especially in classical physics. What rational reconstructionism had no use for was *philosophical* history of science, that is, history of science construed from the point of view of a philosophy of history, specifically in Hegelian or dialectical modes. Such 'history' was regarded as metaphysical, in the worst sense. But even apart from this the *rational reconstructionist* philosophy of science has always been uncomfortable with history generally. It is a messy domain, perhaps the most recalcitrant of the social sciences, in terms of reduction to the manageable frameworks of the hypothetico-deductive or nomological model of explanation. A word needs to be said here in defense of the richer sociological contributions of early logical positivism, e.g. in the works of Edgar Zilsel, of Otto Neurath, and even

of Hahn and Reichenbach. But that involves an assessment of the *earlier* Vienna Circle in contrast to its later epigonic stage – the one we are mainly concerned with.

It is useful to stop here for a moment to review the alternative *ontologies* of science presupposed by these different views, especially since I will argue for a fundamental ontological replacement later.

By an ontology of science, I don't mean a scientific ontology – i.e. an ontology of the domain over which science ranges, but rather the ontology of the entity *science* itself.

The rational reconstructionist ontology may be sketched thus: Science is a *body of knowledge*; or at a given stage, the synchronous time-slice of this body of knowledge; it is constituted, for purposes of characterizing it, by a set of statements related to each other logically, i.e. by rules of inference. These are the acceptable statements: some of them may turn out to be true, others false. But all are entertained as truth-claims or as proferred truths, or as verifiable, or falsifiable – the verificationist, confirmationist, corroborationist, falsificationist issue doesn't concern us here – all are brothers under the skin). The constitution of this set of statements is by means of the agreement of the scientific community as to their acceptability. But the *genesis* of these statements, their history, and the *actual* conditions under which they come to be agreed upon (as against the formulation of ideal or formal rules or conditions under which they *ought* to be agreed on) – all of these lie *outside* and *beyond* the task, (and therefore outside the ontology) of the rational reconstructionist view of science. They are rather held to lie in the domain of the context of discovery. The only statements about which we can say that they are generated *within* the system, are those which are deductive consequences of others. And if they are reconstructed as logical consequences, such statements may be said to already be *contained* in the system, and merely actualized for practical purposes by valid inference. Otherwise they are synthetic statements which may be *added* to the system from the outside, as observation or protocol statements – i.e. those based on subject-neutral or intersubjectivily acceptable experience, according to the canons of the science. So-called statements of initial conditions, measurement statements, and singular statements are of this sort. Although such statements may be weakly construed as 'historical', in that they are not timeless deductive consequence in the system or set

of statements, and insofar as they may be time-indexed, yet they are not historical in any other sense, and their historicity (and their time index) is indifferent to the logic of the system.

In short, within such a rational-reconstructionist ontology of science, the *context of justification-context of discovery* dichotomy leaves only the context of justification as the appropriate domain of the philosophy of science. The ontology of science which is determined thereby is that of an *a*historical, synchronous (and time-independent) body of *achieved* scientific knowledge. The task of philosophy of science is thus taken to be rational reconstruction of this achieved body of knowledge – (or of the set of statements which express it) – ideally, in the form of a logical system, and ultimately, in an axiomatic form.

The ontology of science of the historian of science is sharply different. Here, the *events* of science, in their sequence and interconnection, constitute a diachronic content to be reconstituted from the record (or the *archive*, to use Foucault's term), and includes *both* the theoretical statements or systems, and also the actions, institutions, beliefs, motives and interests of the community of scientists. This sequence of events reconstructed from the record may be construed as a history of scientific theories, and of internal practices (e.g. modes of experiment, instrumentation, measurement). Or it may be construed in a wider sense, to include, as part of the history of science, the economic, political, social, technological, biographical, i.e. 'external' contexts in which scientific theories and 'internal' scientific practises arise and fall.

In this historian's ontology of science, the *logic* of the historically reconstructed sequence may be one of causal relation, either as a simple linear succession of causes and effects, or as a cumulative development, e.g. as the unfolding of reason, or of form, or of God's purpose, or of truth. Or again, the *logic* of the history of science may be conceived dialectically, as one of conflicts and resolutions; or as the progressive revelation and elaboration of an invariant theme, whether *Identity*, or *Economy of Thought*, or *utility*. Each of these, then, may be taken as characterizing an ontology of science, i.e. the kinds of entities, events, processes, causal relations, properties, which constitute the domain of science, and the course and means of its development. In short (in any of the above modes), science may be seen to have a *thematic* history, or one which has a logic of a sort; or it may be reconstructed as sheer

chronicle, allegedly recording nothing more than the uninterpreted facts, or the sequence or temporal order of events. The ontology in any case, is one of events, or of actions, or reasons, or motives, or purposes, but all of these as a diachronic process. Occasionally, the synchronic 'time-slice' 'history' of a given period seems to violate this condition: But even here, the account is not time-indifferent: the temporality of the events, and their sequence, is crucial.

I will argue later that both diachronic and synchronic ontologies are, as such, inadequate; though aspects of both need to be retained in an adequate ontology of science.

To return to our typology:

(3) *Agreeable Internal Relations – Classical:* This type is qualitatively different from the others. Here, by internal relations, I mean the inter-action of philosophy and history of science, in such a way that each, in part at least, constitutes the other, or gives it the characteristics it has. This is where we should end, optimally, for then, we would presumably have an historical philosophy of science and a philosophical history of science. But I speak here only of the *classical* stage, which precedes the *modern* or contemporary one, i.e. of *disagreeable internal* relations. The classical stage includes the classical philosophical histories of science and the classical historical philosophies of science. In fact, the classical stage is classical precisely to the effect that historical and philosophical *divisions* do not yet appear in it. It is like Anaximander's *indeterminate* in which the different elements are contained, but not yet separated out.

In the past, in what might count as a Golden Age, the history of science was itself construed as philosophical history, and indeed as a nor-mative history. Such a history could be constructed as embodying an *a priori* norm of rationality; or, if not, the norm could be discovered and reconstructed by attention to the historical evidence itself. Here, Hegel sets the pattern: The notion of a dialectical history of ideas was first discovered by Hegel as the *modus operandi* of the history of philosophy itself. Though the models are much older – (i.e. of Reason, or God, or the Idea elaborating itself in the vehicle of history, in the actions or thoughts of men, or in the City of Man) – Hegel distinctively constructed the full *logic* of this development (first recognized in the history of philosophy) in his *Phenomenology* (as a science of consciousness), in the History of Nations, and of Society and Law (in his *Philosophy of*

History, and his *Philosophy of Right*) and *also* in his work on the philosophy of nature and of science in which nature itself partakes in this history, and becomes self-conscious in the reflection of this process in theoretical consciousness (in his *Philosophy of Nature*). This is not, properly speaking, history of science as yet, though Hegel studied the history of the sciences and of mathematics for the materials of his *Naturphilosophie*. His interest was systematic rather than simply historical or reconstructive, as such. One might call this sort of construction *tendentious* history of science, in the sense that the history of science was seen as innately philosophical, as the vehicle for the elaboration of the Idea. This *tendentious* history of science was philosophical in the sense that Aristotle in the *Poetics* demarcated *Poetry* as philosophical by contrast to History as mere chronicle (the Ideographic-nomothetic distinction). What was sought, in history, were the evidences of a universal (or if one prefers, of an invariance). History was in the service of a Universal Idea,

1. whether the Idea was that of Reason unfolding itself by critical negation (Hegelian 'falsificationism' – which would lead one to characterize Popper as a crypto-Hegelian in a mode of psychological abreaction; and Lakatos as a closet-Hegelian, because *he* knew he was one, but didn't want anyone else to. But more on this later);

2. or whether the Idea was that of eliminative positivism, in which the history of science was a history of the overcoming and elimination of dogma, superstition and metaphysics by 'positive fact';

3. or whether the Idea was 'Economy of Thought' as the adaptive, biological form of a principle of least action as a strategy of cognitive success (e.g. as proposed by Ernst Mach);

4. or whether this idea took the form of the reduction of all phenomena to the intelligible or rational form of causal law, or of some overarching Principle of Identity (which Myerson saw revealed in the history of physical science).

In all these cases, and in others besides (e.g. in Poincaré, with *invariance*; in Peirce, with asymptotic approximation to reality, and increasing determinateness), there is seen to be some general or universal heuristic or methodological principle by means of which the history of science is organized, and through which it can come to be understood in its development. Poincaré and Duhem had both insisted on the essential role of

theoretical *preconceptions* within science itself (without which, as Poincaré had said, all experiments would be sterile; und upon which, as Duhem had pointed out, experiments so much depended). This same approach to the history of science was characteristic of the most fruitful and epochal historical works of this period (e.g. Mach's *Principles of Mechanics* and *Root and History of the Principle of the Conservation of Energy*; Meyerson's *Cheminement de la Pensée*, and *Identité et Realité*; Duhem's *Aim and Structure of Physical Theory*). The methodological principle of starting with a priori assumptions within science itself had already been proposed by Kant in two works: The *Metaphysische Anfangsgrunde der Naturwissenschaft* and *Von Übergnag von den Metaphysischen Anfangsgrunde der Wissenschaft zur Physik*. Why not, then, take this same approach to the study of the *history* of science?

Thus, in fact, with such names as Whewell, Mach, Duhem, Poincaré, Meyerson, Lalande, Brunschvicq we pass over from what might be construed as purely *a priori* reconstructions of science and the history of science in the service of an Idea, in Kant and Hegel, to a modified philosophical history of science and historical philosophy of science as two interweaving aspects of a *single* discipline. The internal relations of history and philosophy of science were not only agreeable, in this classical age; they were essential since the single unified discipline was constituted by this interrelationship.

(4) *Disagreeable Internal Relations – The Modern Period:* The idyllic classic age or Golden Age of the philosophical historians of science was followed by an unravelling or separation of the elements already contained in their thought. Mach was the great inspiration of the logical positivist Vienna Circle (i.e. the *Mach Gesellschaft*), and his methodological dicta concerning reduction of all theoretical formulas to sensory elements became a systematic basis for the new logical construction or reconstruction of the sciences (*'Logische Aufbau'*) undertaken by Carnap and his followers. Duhem's so-called conventionalism, first developed by him as a solution to the problem of theoretical change in the history of science, passed over into a pragmatic relativism with respect to linguistic (and therefore also ontological) frameworks which we find in Carnap and Quine. These systematic concerns with the logical and linguistic reconstruction of science constituted a distinct break with the historical concerns of the earlier generation.

Lenin was among the first to launch a full scale realist attack upon what he saw as the epistemological consequence – (i.e. phenomenalist idealism) – of the new empiricist positivism which derived from Mach (in Lenin's attack upon Bazarov, Bogdanov, Yushkevich and others of the Russian school of empirio-critics, and upon Poincaré, Duhem, Karl Pearson and the young Philip Frank, among others – in *Materialism and Empirio Criticism* (1908)). But the *modern* strife between history and philosophy of science did not reemerge until some philosophers, in collusion with historians of science, provided the historians themselves with a *philosophical* rationale for their attack on positivism.

A most important early work here was E. A. Burtt's *The Metaphysical Foundations of Modern Science*, (1924, revised 1932). In this work, that very metaphysics which the positivist program (in those *same* years) was trying to eradicate and eliminate from rationally respectable science as a hangover from its *pre*-scientific stage, was placed instead at the very foundations of modern science – i.e. at its *historical* origins, in Copernicus, Kepler, Galileo, Descartes, Gilbert, Boyle and Newton. Nor was it ascribed to the 'backwardness' of these minds, as a remnant of their prescientific training. It was seen as essential to the theory-formation of classical physics itself. Duhem tried to push metaphysics out the front door (in *The Aim and Structure of Physical Theory*) but immediately invited it in the back door, splitting himself into positivist-scientist on the one hand, and 'believer' on the other. Popper, in his ambivalent campaign for a realist epistemology, also permitted a pale and anemic ontologism in through the back door: after first demarcating science from metaphysics, he then proceeded to find a role for metaphysics in the *critical* dialectic of scientific thought. His students and followers – Feyerabend, Agassi, Lakatos, Watkins pushed the door open further. And finally, because of their interest and work in the history of science, their increasing rejection of the demarcation criterion between science and metaphysics, and their sharpening anti-positivism and anti-sensationalism (anti-inductivist-empiricist), the influence of these followers and students of Popper upon the historians grew. The Popperians, ex-Popperians and post-Popperians at least took history of science and historiography of science seriously. (The two crucial works here are Lakatos' *Proofs and Refutations*, dealing with a case study in the history of mathematics, and Agassi's *Toward a Historiography of Science*).

Popper's effects were enhanced by those of Polanyi – himself a scientist of note, turned philosopher-critic-historian and methodologist of science. His strong anti-positivism, and his insistence on the personal and tacit dimensions of scientific thought – in discovery, and in the understanding of science – also fed the fires of the anti-positivist, pro-historicist mood of some historians of science. Also, two philosophers – both in one or another fashion historically oriented, both rejecting the received wisdom of logical reconstruction, and both strongly influenced by Wittgenstein – added insult to injury by resuscitating the contexts of discovery and rejecting the straitjacket of the context of justification. Toulmin, in *Foresight and Understanding* and Hanson, in *Patterns of Discovery*, both recovered themes which had been earlier rejected as psychological, and therefore beyond the pale.

There is more history to this story, but its happy and serendipitous culmination came when a bona-fide historian, Thomas Kuhn, with an excellent book on Copernicus to his credit, and with the influences of his colleagues – Leonard Nash and Paul Feyerabend, and of Polanyi as well – came out with a *blockbuster* – *The Structure of Scientific Revolutions* – an irritating, naive, confused and provocative work, which nevertheless, brought it all together in one glorious explosion.

The sociology of the reception of *The Structure of Scientific Revolutions* is an interesting subject in its own right, and I won't try to get into it here. But the results are well known. Anti-positivist, anti-reconstruc-tionist *history* of science had found a naive but provocative formulation which shook up the philosophers of science. Small wonder. They knew little history – especially history of philosophy – so it all looked new to them. They were incompetent sociologists, so the sociology of science alleged in Kuhn's work simply had to be accepted enthusiastically or rejected violently on a priori philosophical grounds. What the philos-ophers *could* do was talk about paradigms (which they did endlessly). But the effect, one way or another, was to turn many of them to the history of science, and to afford a new legitimation for the philosophical reflection on this history, whether in criticism of or in agreement with Kuhn.

Well, here we are, with a Disagreeable Internal Relation between philosophy of science and history of science – disagreeable because the old *paradigm* – (excuse me, I mean *model*) – of inherently rational science

to be explicitly rationally reconstructed, *and* the old model of the subject-neutral and firm reduction base in observation had been eroded from *within* philosophy, and on the grounds of an *historical* reconstruction of the growth of science.

The typology thus far is missing one component: the *Modern Stage* of *Agreeable Internal Relations*. But this will require a formulation of the internal relations between history and philosophy of science which is more than simply an account of the relations between two disciplines – and also more than the next stage in this schematic dialectic of their relations. It will require a more systematic *ontological* reconstruction of the entity called *science*, i.e., a better philosophy of science; so that the relation between the study of its history, and the study of its epistemological and logical character will flow from that reconstructed ontology itself. All I can give here is a programmatic sketch of a larger enterprise – if you like, a research program – and one which I noted briefly at the outset in setting forth my own theses.

<center>II</center>

Perhaps I should begin by rejecting two close approximations to the program: Kuhn's and Lakatos'.

Kuhn characterizes science as a social enterprise, constituted by men with beliefs, needs, ambitions; but all institutionalized in their modes of action as scientists – i.e. constituting a scientific community. The shifts and changes in this enterprise are seen as changes of a human sort – not only rational, but social-psychological as well. All of this is set into the context of *stages* of scientific growth, each with its investment in a technology – i.e. not only in the usual sense, but also in the sense of an articulated cognitive and practical technology of rules, procedures, accepted truths, and modes of conceptualisation and of practice. Therefore each such stage, i.e. so-called normal science, exhibits a built-in *tenacity* to remain unchanged. *Proliferation* [I borrow the tenacity and proliferation vocabulary from Feyerabend, who borrows it, in part, from Peirce] comes from within the community, but for reasons not yet clear. In the original version in *The Structure of Scientific Revolutions*, Kuhn has the shift take place when anomalies are obvious in the old framework – but not because they are. In this sense, he rejects the

dialectic of critical negation as too formalist and rationalist a model of what *actually* goes on, which he recognizes as much messier. But in place of this 'clean' rational dialectic, Kuhn resorts to the irrational and nonrational as the motive power of historical changes in theory. Thus, he speaks of 'leaps of faith', and quotes Planck's famous remark about a theory dying only when its proponents die. There is a teasing authenticity to this humanized and socialized account of science. It gives a description of the historical sequence, and even suggests that since that *is* how science *did* grow, that's how it *should* grow. *Is* converts to *ought*, by the *success*-transformation. So Kuhn can say (in his 'Reply to Critics', in *Criticism and the Growth of Knowledge*) that his is not just a value-free descriptivist sociologism, but is, 'in a sense', also prescriptive.

Lakatos, in his merry way, had in the meantime cooked up a methodology for science which is frankly normative, by (his) contrast to Kuhn's (which he criticized as non-normative). Rejecting Popper$_1$ .... Popper$_n$, Lakatos wants to find, in the rational dialectic of criticism within scientific theory, that invariance which demarcates good from bad science. His unit is not the theory, and his criterion is neither truth nor falsifiability. It is, rather, whatever advances scientific research, broadens it, solves problems better than what preceded it. The unit is not the theory, but the program of research (more complex, but no clearer than the paradigm. The phrase 'research-programmes' is Agassi's originally, in this context.) So some of these are *progressive*, some *degenerating*. Not everything goes. And since *all* theories are *born* false – (a straight-out Hegelian view, since all theories are incomplete, and if thus inherently fallible then false – no theories are privileged in this way *except* insofar as they *increase* knowledge. Dialectic there is – but history as such provides no norms. Only the principle of selection, within the history of science, of those programs which are *progressive* or which shift the cognitive activity of science in a progressive direction by leading to *new* knowledge – only such a selection principle yields the *normative* dialectic of good science, and hence, of rationality itself, in the example.

I cannot pretend to cope with these two views here, except to caricature them both: With Kuhn, we get history without dialectic; with Lakatos we get dialectic without history (by the use of history

as no more than the source of examples of good science, or of the norm of science). The mediation is obvious. And also obvious is the requirement that we transcend these two views, and not merely combine them. What we need therefore is a dialectical history of science. But having said this, it should be clear that the norms of such a science must be derived from the historical context itself and cannot be imposed upon it by philosophical fiat, or by some supra-historical *a priori* conception of rationality. This can't be done by finding an invariance in the history, by some abstractive or inductive means. It requires, I think, a characterization of the *historical enterprise* called science, in a new way, in order to see in it the sources of its *intrinsic* normativeness.

I have to be very brief and elementary here. I start with the characterization of science as a mode of cognitive human action, or praxis. Like all human action or praxis, this activity is teleological. By this, I mean not only that science serves some function – for so do the lungs, and so do leaves on a tree, and that's not what I mean by teleology – but that it is *intended* to serve some purpose. That is, it is undertaken by conscious agents in order to serve some purpose. It is, in short, intentional activity. Moreover, like every mode of human action, it is social. That is to say, its context is social human existence, and its teleology is a social teleology. The needs it serves, and its conscious purposes in serving these needs, are social needs, and socially conscious purposes. As *cognitive* praxis, it serves these needs by means of the activity of acquiring knowledge. Science, therefore, is a practice of acquiring knowledge, whose presupposition is that such an acquisition of knowledge itself serves a social human need or purpose.

Now there are other modes of cognitive praxis – in fact, no form of *human* praxis is non-cognitive. For example, sheer skill or craft, or knowing how to proceed, is cognitive in this sense. So too is all ordinary human action. I would argue that all the forms of social human interaction, from interpersonal to social and political activity, to theoretical activity, are cognitive, involving either language, or some other symbolic mode of communication, and requiring the socialization and learning of the rules and concepts of one's peers, one's parents, one's social group, and one's profession.

What is distinctive of science, as a mode of cognitive praxis, however, is that it involves *theory*. That is to say, what science acquires as

knowledge is neither skill nor craft nor social rules, though the imprint of all of these is upon the scientific mode of cognition. What it acquires as knowledge is *theoretical* knowledge. *Science is thus the theoretical mode of cognitive human praxis.* The history of science is therefore the history of this distinctive mode of cognitive activity and of the conditions of its genesis, growth and change.

But what is theory? Here I want to be perversely brief, and present my thesis, but not yet its argument: Theory is the mode of knowledge whose object or content is truth (and *not* convenience, utility, pleasure, or rightness, etc.). The very formation of the concept of truth is the genesis of theoretical knowledge. To know a skill, or a craft, or to know how to get on with others, is *not* a matter of knowing what is true. Such cognitive activity is instrumental in its own way, at its own level. But social development of the species introduces a distinctive value in human life, not present in other species as far as I can understand, namely, that of reflectively formulating assertions whose value (in both senses: as *survival* value, and as propositional value) is truth; or to put it differently, assertions which may be true or false, i.e. proferred truths, or truth-claims.

Part of the program of an historical epistemology is to account for the genesis of this feature of human cognitive praxis, not simply as a biological feature of adaptation, but as the product of human social evolution – i.e. as a postbiological or *historical* achievement. The behaviorist reduction of truth to a matching-function between two states (one a propositional state or a verbal behavior and the other a state of the environment), simply won't do here, though the instinct for correspondence is sound. The relation is much more complex. But suffice it to say here, sketchily, that only when an acquired cognitive achievement – e.g. a skill, a craft, a mode of action – can be detached, so to speak, from its actual performance, and reconstructed or represented symbolically, in a language, can the cognition *begin* to approach the conditions of a truth-claim, i.e. the condition of theory.

In linking the genesis of theory to the formulation of truth-claims, i.e. of assertions. I also want to make an instrumentalist claim without proposing an instrumentalist theory of truth: Cognition is instrumental in serving human needs and purposes. Its warrant is its social utility. But cognition would have no utility, if *truth* had no utility, i.e. if theoretical

knowledge had no utility. The condition for the genesis of theory as a mode of cognitive praxis is the practicality of theory in human life, both as an instrument for the satisfaction of the needs of human existence – i.e. for the mastery of nature as a resource for human life – and for the purposes of social organization of human activity for these purposes. An older formulation puts it thus: It is not that what is practical is true, but rather, that what is true is practical. But that is in fact too pat a credo. Rather, let me reformulate it thus: that theory, as the cognitive mode of the search for truth, is practical. Looking for the truth is what is practical and efficacious about science. Finding it is the hope of science. Science may then be characterized as rational hopefulness: Not empty wish, but deliberate activity or practice in the service of realizing such a hope.

What I am proposing here, only programatically and sketchily, is what I would characterize as an historical-materialist theory of the genesis of theory, or of theoretical cognitive praxis. The apparent instrumentalism of truth in such a view, as I set forth here, is then replaced by a realist emphasis: As a representation in some symbolic form of a mode of action or practice, the theoretical formation has its truth-value in the adequacy of the representation: i.e. in its practical conformity to a successful mode of activity, which it institutionalizes, so to speak, in the representation. It is *not*, therefore, an unmediated representation of some external state of affairs, but rather one which is mediated by the practice or mode of action which it represents. The sheer externality of a state of affairs becomes 'objective' for us, then, only as it is mediated by our practice. What we can know is therefore always conditioned by the way that we come to know it. In a sense, our knowledge of the 'external' world is a knowledge of what this externality is amenable to, in our incursions upon it and intervention in it.

An action itself succeeds or not – or more complexly, a process of action, or a rule of action or a strategy – succeeds or not, depending on how well it realizes the external (in this sense, objective) constraints of the world in which the operation or action proceeds. That world, in turn, is independent of the action. The action doesn't construct it; it reconstructs or transforms. It is in the activity of transforming an external world to human use, by means of action or praxis, that practical knowledge of that world is acquired. It is by means of the *representation* of successful action that theoretical knowledge of that world is acquired.

Theory, then, is that form of action which is mediated by representation in a language, i.e. symbolically. Or, to put it in a much more traditional form, theory is reflective practice – or the very practice of reflecting on practice. In full dialectical jargon, one would say that theory is practice become self-reflective. All human practice or action is already cognitive, though not yet theoretical. It is a means of acquiring knowledge of the world by transforming it to meet needs, purposes, interests. This transformation, in turn, is a necessary life-activity – it is the life activity of the species, its mode of reproducing its existence. But such cognitive activity achieves a distinctive human form when it becomes reflective: When it can be embodied in a representation, communicated, transmitted as acquired knowledge from one generation to the next – in short, when it becomes theoretical. In this sense, humans are distinctively theoretical animals.

To begin in this way with the genesis of science is ambitious, if not arrogant. But if one is to build into the conception of science both its teleology and its normativeness, on the one hand, and also its historicity on the other, these features must be built into the genesis of science itself. Please forgive, therefore, the briefness and perhaps the density of this sketch. It is at best, programmatic here. Still, certain conclusions follow from it, in this genetic ontology of science:

First, that theorizing is a natural requirement of human action, and not an extraneous one.

Second, that it proceeds from and depends on cognitive praxis which is not yet theoretical – but then, it comes to inform this praxis itself, as conscious expectation in the imagination becomes a distinctive feature of human practice, and the intentionality of human action is achieved in this reflective and teleological mode.

Third, that *all* modes of such theoretical praxis are already included in what I would call science and that science, in this broad sense, is of much greater antiquity than standard histories or philosophies of science allow (unless for purpose of historical demarcation, one wants to call such early, yet *already* theoretical modes of cognition *proto-scientific*).

Fourth, that the normativeness of such theoretical practice is built into its teleology, It is intrinsically teleological; its norm, and its practical utility, its *truth* (thus far inadequately specified here);

Fifth, and most crucial for our purposes: such theoretical cognitive activity or *praxis* changes and develops with the changes in the needs and purposes it serves; and may be cumulative, only insofar as there are perduring historical needs and purposes which have to be met at *all* times. Its dynamic character – its proliferation, to use Feyerabend's term – is a function of the historical evolution of human needs, as societies change and grow; its stable character – its tenacity and its cumulativeness – is a function of the successive achievements of theory with respect to meeting transhistorical (but *not* suprahistorical) needs – in particular, those associated with the understanding of the natural world as the abiding resource of human existence, the permanent environment of our species-existence. The context of theory-change is therefore the context of *historical* change of the forms of life of the species. Since I take this historical change itself to be a dynamic of social forms of the organization of the production of the means of existence and of the reproduction of the species, in historical materialist terms, the history of theoretical change is necessarily interwoven with this history: *not* as its reflex, in some simple-minded base-superstructure theory, or in the vulgar Marxism of simple copy-or reflection-theory but rather, insofar as theory itself is a vital and efficacious component of this history.

But what of *scientific* theory? Is all theorizing scientific? I admitted, earlier, a useful demarcation between proto-science and science. I think the demarcation comes, historically, with the division of labor. When theoretical cognitive praxis becomes a distinctive social institution – i.e. when the role of theorizing becomes the function of a *part* of society, then science *as* a social institution begins its truly revolutionary course. In the *Metaphysics*, Aristotle already suggested this division of labor as the beginning of science in his remarks on the origins of rational mathematical theory in Egypt. He suggested that only with a surplus, such that leisure could be provided for a priestly class, could science proper begin. I think he was right.

I also think that this marks science, as an institution with a class-character, and, in such class-societies, with a class role. This complicates its history. It is no longer simply the truth-seeking theoretical activity which represents and also helps to guide successful practice, in some simple sense of an undifferentiated human community providing for

its needs. It becomes involved with its role not simply in the acquisition of theoretical knowledge, but the acquisition of this knowledge for differentiated purposes, in historical class societies. Ideology and special interests intrude on science, in the very conditions of division of labor which give it birth. Socially differentiated needs – *historical* needs – both mediate and distort its truth-acquiring function. The simple progress of rationality is replaced by the messier dialectical process of rationality for some purpose. The point is that what we characterize as rationality, *itself* has a history, and it is a social history.

What I am suggesting, sketchily and programmatically here, is a much richer and more complex ontology of science, and therefore a much more difficult project for the philosophy and history of science. But I am further suggesting that such a philosophy and history of science – whatever the necessary division of labor between philosophers and historians, whatever the requirements of specialization and rigor – must nevertheless be an integrated philosophy and history of science – a philosophical-historical theory of science as a changing and developing social human activity. This is what I would propose, therefore, as the missing type of Agreeable Internal Relations between philosophy and history of science.

Boston University

TELOS AND TECHNIQUE: MODELS AS MODES OF ACTION

[1968]

There is something hopelessly desperate about attempting to find a radically new departure for talking or thinking about the future. At times, it seems a phantasmagorical task, fit for Sisyphus, of finding the key for a lock whose shape is itself vague and indeterminate. We are burdened with the consciousness of our vast achivements, of our apparently limitless technological prospects, of the grim optimism assuring us that whatever we undertake will somehow and at some time be effected. Together with this, we are haunted by the history of moral and political failure, by that vision of the servant become master or of a technology, blind and soulless, which works us to its own dialectic processes and whose rationales we are expedient enough to parade as if these were our own choices and our own ideologies.

So our *Angst* is compounded, on the one hand, of a power and dominion that commits us to what it *can* achieve, because what it can achieve is just what we had intended it to achieve – the control of a recalcitrant nature, to fit it to our needs – and, on the other, of the conviction that we are thus somehow the captives of our own past intentions. Thus our ingenuity, our creativity, the openness of free invention, all seem bound to what we once thought – or what was thought for us – to constitute rational goals and moral ends.

Now I do not propose to moralize about the dead hand of the past or about the limitations on our creative invention of the future, for at its worst this degenerates into pious confessions of impotence or the whining assertion that we are lost, alone, and fallen from the grace of an earlier golden age.

Rather, I would like to attempt something modestly hopeful and not at all desperate – this is, to characterize the normal process of creating the future, by bringing it down from the scale of cosmic crisis to that of daily and local necessity. I do not think there is a millennial solution to the future; but I do think that the pattern of our ordinary planning and striving prefigures whatever larger structures there are in terms of which long-range creation of the future can take place.

Perhaps some truisms may not be amiss here. The first is that man creates his future in the present; the second is that the present is more than a simple summation of the past, for it includes, at least at the level of conscious human activity, the element of anticipation. The epistemological aspect of this resides in our capacity to envision the future.

Unfortunately, in the context of much of epistemology, including that common-sense kind which traditionally conceives of thinking as an intellectual activity 'in the mind,' so to speak, envisioning the future comes to mean thinking about it or imagining it in some private mental space. Since, clearly, it cannot exist in the present as much, since it is the future, the future is assigned a merely ideational status, as an object of thought or imagination. I would argue that this is a false view and that adherence to it leads to passivity and practical impotence, because the future becomes the toy of imagination, the fantasy life of the inadequate present, the realm of unenacted wish, or if enacted at all, then only in imagination.

That this is false even of what is purely intellectual about thinking of the future is something that such diverse thinkers as Aristotle, Descartes, Kant, Marx, Peirce, and Dewey have shown. For them, in various interpretations, thinking is an activity that results in action in the world. Moreover, it is broadly conceived of as that activity which is unique to human beings and is characterized by conscious purpose. Therefore, it is not simply some object held up before the mind within the confines of one's own thoughts.

The prototype of such future-directed action – in which the future is more than the blindly inevitable fact of succession in time and includes some envisioned goal as its content – I would call a model.

Now this is surely perverse, for 'models' in our ordinary sense are imitations, diagrams, scale versions, or pictures of something already existing. However, they can be more than this, as in prototypes, plans, hypothetical constructions of various sorts which serve as guides to action. But I mean more than this by 'model.' Initially, let me suggest that models are the highly specialized part of our technological equipment whose specific function it is to create the future. I cannot mean by model anything quite as narrow as either an imitative version of something already existing, as in scale models, or simply a prototype or plan for some future embodiment. At best, these are what models may look like but not what they function as. To stretch the term 'model' even

further, let me suggest that what I mean by model is not simply the *entity* we take as a model but rather the *mode of action* that such an entity itself represents. In this sense, models are embodiments of purpose and, at the same time, instruments for carrying out such purposes.

Let me attempt to clarify this idea. No entity is a model of anything simply by virtue of looking like, or being like, that thing. Anything is like anything else in an infinite number of respects and certainly in some specifiable respect; thus, if I like, I may take anything as a model of anything else, as long as I can specify the respect in which I take it. There is no restriction on this. Thus an array of teacups, for example, may be taken as a model of the deployment of infantry battalions, and matchsticks as models of mu – mesons, there being some properties that any of these things share with the others. But when we choose something to be a model, we choose it with some end in view, even if that end in view is simply to aid the imagination or the understanding. In the most trivial cases, then, the model is already normative and telic. It is normative in that it is chosen to represent abstractively only certain features of the thing we model, not everything all at once, but those features taken to be important or significant or valuable. The model is telic in that such significance and value can exist only with respect to some end in view or purpose that the model serves.

Let us go one step further. Even in the weakest sense, the model is an aid to the achievement of something in the future which is as yet unachieved. If this is no more than intellectual enlightenment, or a clear understanding, then it is no small gain if we value these as goods. So even in the case where we say the model is only an aid to the imagination or that it only helps us to envision or picture or understand what is happening – for example, when we make a schematic model of some complex organism, or make a prototypical model of something we intend to produce – this in itself is an achievement. In a sense, it is a creation of something working toward the future, something which, at the point of making the model or contemplating it, has not yet been achieved. Of course, that is a very weak sense. But now let us turn to the stronger sense in which the model is an achievement of something in the future. I have claimed that the models are not merely part of the technology for creating the future, that is, they are not merely instruments to be used but are themselves modes of action that, in fact, create the future.

Perhaps, then, we had better unpack the metaphorical term 'creating the future,' or 'inventing the future.' We mean by it, I think, *acting* in

such a way as to make the future conform to some present vision of it. In a trivial and uninteresting sense, any action whatever 'creates the future.' If I raise my hand now, I have changed, in some way, what the future will be. However, 'creation' and, still more, 'invention' carry the connotation of conscious and purposive action (if not also of rational action), as well as that sort of conscious and purposive action which has the highest values set upon it, whatever these are taken to be.

Now let me turn Kantian for a moment. That action or mode of action which I take the model to be is a model action. If I may be forgiven the pun here, I mean to say that it is a model action in the sense that it sets up a normative prototype whose function is not simply descriptive but imperative. It is an action or mode of action that one could rationally will to be universally emulated. Thus, it is more than an action; it is at the same time a *call* to action. It functions (in the way of this paraphrase of the categorical imperative of Kant) as an exemplary model with a claim on us to share in its purpose and in its mode of action to achieve this purpose.

What seems queer here is that our usual wire and paper-clip models or our plaster mock-ups seem more modest than this. It is hard to attach this resounding phrase 'call to action' to something as simple, for example, as a scale model or some string-and-wire model. But this is a matter of scale and scope rather than of function. The values of a proper construction, in architecture or in logic, seem small compared to that larger human future which our large-scale social and technological models envision. But the demand of the normative construction of a logical proof or of model building is still: "This is how it ought to be done; this is what needs to be understood; this is how one ought to operate." Thus alternative models, as embodiments of purpose, present choices among alternative purposes. But, furthermore, models as embodiments of one or another kind of belief are, as C. S. Peirce would have pointed out, exemplifications of readiness to act in a certain way. To believe something, according to Peirce, is to be prepared to act in such and such a way. Therefore, we may say that the belief is already part of the action, and not something that stands 'behind' the action and 'waits' for the action to take place. The action, as it is considered here, is not just something that begins at point A and ends at point B but is rather something already embodied in the tacit or explicit 'beliefs' that the model represents, which precede the overt activity itself.

Authors of ideal imitation theories in aesthetics, drawing on

Aristotle's profound ideas, talk of *emulating* the divine way or Nature's way of working – not copying God's or Nature's handiwork, but rather emulating the mode of operation, of which the handiwork is only the exemplification. The master stands in the same relation to the apprentice: the teacher is the model not simply of this or that but of a *way* of work, a technique of inquiry, a kind of commitment. The emphasis is on the mode of action which the model represents, that is, a mode of action with respect to some end in view. So it is two-sided: on the one hand, it is an embodiment of a purpose; on the other hand, it is also, dynamically, the means of realization of the purpose. The purposes here are, at the same time, *entertained and undertaken* in the model.

So far, I have said that a model is an aid to the understanding or to the imagination. It is also a call to action in the sense of presenting some exemplary form of *how* things ought to be done or *what* ought to be done. This normative claim of the model is strongest when the model goes beyond the exemplary function. Even in the exemplary function of models the future is to be patterned after some present version, even if this is presently an ideal version. But in being 'ideal,' this present version is taken to be good and, therefore, bears the claim that one *ought* to emulate it. The limits of conception here delimit the present ideal. But some models do more than serve in the creation of a future in terms of some present ideal; they transform the total vision of those who involve themselves in the model. They are radical and revolutionary in their effect. They are modes of action which act by infection upon an age and its consciousness, and they help release energies and possibilities of the imagination and mind beyond anything that the model itself exemplifies. That is, the model produces more than it contains. It is a nonconservative model in this sense. It is perhaps hard to think of what such a model would be like unless one gives some examples, which broaden the concept of 'model' still further.

I think poetry, at its best, works such a transformation of vision. Similarly, ethical and religious and political ideas have triggered such effects. That is, they have produced more than they were taken to contain in any dated version or interpretation of them, as have the not-so-exemplary lives of great men – for often the life of a great man may be all but exemplary in the ordinary sense. The point of emulating a great man is not merely to live as he lived but to create as he created – not to emulate his life in any other respect than to be able to achieve the kind of creativity, or the mode of activity, which he represents.

The purposes that such a model presents, the goals and ideals to which it summons commitment, are not simply 'in it,' as lessons to be learned or as raisins in a cake. They are read out of it by the interpretive and creative genius, the inventiveness of those who respond to it. The model thus generates creative action. How it does this, I think, has not been investigated and is not clearly known.

If one were to examine the theological and religious version of this sort of model, it would be instructive. We are told that 'God is made great by belief,' which we may choose to translate (after the fashion of Lugwig Feuerbach) as meaning that 'God's greatness is nothing but the greatness of the belief in God.' We may paraphrase *this* paraphrase in our own context: 'The greatness of the future is nothing but the greatness of our vision of the future.' There is nothing 'great' about the future. In a sense it doesn't have any character at all which we can ascertain, except in our present vision of it. The future may be very different from our present models of it. But it will not come about randomly. If it is a human future at all, if human actions have some part in shaping it, it will be different from the present models because different and alternative models have been created, each in its way coming to constitute the future in the ongoing process of 'present' actions. There is no other place where real work is done, except in the present, and work without a model is brute work, not action at all in the sense of consciously purposed activity; it is not human action at all but at best a dehumanized and brute activity. What we should thus characterize as distinctively human action is that technique which is governed by *telos*, by conscious purpose; and a 'model' in this broad sense, is the embodiment of that kind of action, that is, the embodiment of creative human purpose.

In this sense, the creative model is not simply one that exemplifies the creativity of whoever proposes or enacts it 'Creative' here should be taken to apply to a model that calls forth the creativity of men, releases their energies from the present, and serves to transform it radically.

The features of such a model are, then, that it is necessarily critical of the present, and not simply an envisioning of the future. Thus, such a model has to be both a vision and a creation of the future that is not simply 'more of the same,' or that simply represents the present at a later time. The future here has to be one that destroys the present, preserving only those elements of it which are future-now, those which represent radical innovation in the present.

There are therefore things that are *presently* 'future,' to use this

paradoxical locution. Since they ought to be preserved, they are not 'obsolescibles.' There are things that are 'past-now,' which are 'obsolescibles.' Which to preserve and which to eliminate is a choice but not a choice that is made merely in theory. It is a choice which has to be made in practice. But even this is a false division because the theoretical choice, if it is a real choice, is one that will terminate in an action of some sort, or is already the inception of an action of some sort. In this context, the broad sense in which I am using the term 'model' is one involving such choices and actions, whether these choices or actions are embodied in some visible or tangible form or whether they are in the more fluid form that Mr. Ozbekhan discussed previously. Parenthetically, I should say that it struck me that there is a certain conformity between this conception of the model and Mr. Ozbekhan's analysis of so-called operational models, strategic ones and normative ones. He called them 'plans,' but I think we mean the same thing.

What I had intended to call *technical* models are what Ozbekhan calls operational and strategic ones. What I call an *ideal* model in the exemplary sense is the normative model – that is, what we take to be a proposive model of what ought to be. But beyond this, there is needed something like the last conception of model which I propose: a revolutionary model, which goes beyond the presentation of norms and which is creative and critical in the sense of a criticism of 'oughts,' a criticism of these norms themselves. Such a model may be embodied in a number of forms.

It has been mentioned here that revolutionary political parties may represent such a 'model.' I shall add that there are all sorts of revolutionary models in the world; some of them are political parties, some of them are works of art, some are environments; some of them are persons and lives lived.[1] Each of these presents here what I regard as a revolutionary proposal: not simply 'This is how these *specific* things ought to be done,' but 'This is how one ought to approach the doing of things; this is the mode of operation one ought to adopt.'

Because of the abstractness of this presentation, I ought to give some examples, which flow from our present experience. Perhaps one example of this extended sense of 'model' would be this conference itself. It is a mode of action. Simply described, of course, it is a conference of people called for such and such a purpose. Also, it is a model in the sense of proposing some exemplar of how things ought to be done. Whether we accept this exemplary model or not is a matter of criticism and evalution.

But it proposes that this is how, in this particular case, one ought to go about talking about or inventing the future environment.

There are, clearly, alternative ways of going about this same task, and either one may propose, 'Let's get the best man and let him decide and go ahead' or one may say, 'After all, we are architects, and though it may be enlightening to hear what others have to say, they are not addressing themselves to 'our thing,' and their problems are not the same as ours.' There are, thus, alternative ways in which this meeting could have been called or this purpose served. These are the alternative models. But *criticism* of the alternatives is something that goes beyond the actual modes of action involved in such alternatives. Rather it takes such alternatives as objects of criticism; that is, it is critical reflection on action or on alternative modes of action.

Professor Tillinghast's paper presented alternative models of work; that is, the 'work' model was used to represent other things. Sex was presented under the 'work' model. Leisure was presented under the 'work' model. Here, then, is a particular way of envisioning these activities. In the peculiar case here, a model was presented as something to arouse our critical opposition, to reveal to us that when things are presented under this aspect, they are clearly not acceptable. That is a normative approach, somewhat akin to the art of caricature, which focuses on features and qualities of physiognomy that give striking insight into character and appearance.

There is one other homely example that I would like to take from the discussion following Fein's paper. It is interesting to me that architects talk about their 'clients.' Think for a moment about the professions that take clients (such as law and architecture). By contrast, medical practice takes 'patients.' Much is revealed in this difference. A patient is someone who is *passive*, upon whom an agent acts. The psychiatrist or clinical psychologist talks about his 'patient' as a 'client' sometimes, as in the phrase 'client-centered therapy.' What is the difference? A *client* presumably has something to say about what he wants or needs; a *patient* does not. Most doctors are upset by patients who act like clients, since, as everyone knows, 'The doctor knows best.' What I am suggesting is that these are *alternative* models *as modes of action.* Each entails some complex union of *telos* and technique, some distinctive projection of ends and choice of means, that we may represent abstractly in a conceptual model.

In all of these cases what I am proposing is that the model here is not

simply our way of entertaining a hypothesis or of bringing things before our view in order to consider alternatives. This is rather too passive an account of the interaction of ideas and action. The model is rather a tentative commitment, already the initiation of an action. The function of a model is not simply to entertain an action but to undertake it. Or to put it differently, to entertain an action is, in this sense, already to undertake it in some degree.

The argument here is that models are more than abstract ideas. They are technological means for conceptual exploration leading to experimentation. But an experiment is something that has to be performed and not merely conceived to be useful. In this sense, models are experimental probes, essential parts of the human technique for confronting the future – but not as a passive encounter with something already formed. Rather, in that unique way in which human action is creative, such an encounter shapes the future. Thus, we may suggest that models constitute the distinctive technology of purpose.

DISCUSSION

Tillinghast: You said the future has destroyed the present, that other things are obsolete already. How do you react to these without having seen the future?

Wartofsky: Either you force the future to conform to your model of it, or you don't do anything about it.

Tillinghast: I thought you were against forcing the future on your model.

Wartofsky: What I am saying is we shouldn't try to force *more of the same*. Any model which is a model of what presently exists is not a telic model at all. It is simply a duplicate of what already exists. It has no normative content whatsoever. A model has to pick out certain things as being important relevant characteristics. If it is not selective, in terms of relevance, it is not a model. I can certainly re-create the present in the future, but that is an empty sense of 'future.' In the sense in which I am using the notion of model here, it entails that kind of future which is the outcome of combined *telos* and *technique*, which is an embodiment of some purpose. If the future-directed purpose is to retain the present – and if one can find good arguments for this – then I would have to go along with the notion that such a future is simply more of the present. But I can't see that this option is serious. I can't even think of a

period in history when this would have been true. But certainly even if it were so (supposing there were people who believed that the future is already here, that the millennium has arrived), there would always be others whose vision of the future required abolition of the present, or of specific features of it.

Mazlish: Let me comment on the last point you made. Aren't you running roughshod with experimentation in human affairs, as distinct from experimentation in the natural sciences, when you talk about the proliferation of the former? You are talking about the ongoing system, and the problem is that you can do one thing, but you are not going to be able to go back and do the other. My first question is: Is there more to this than what you said?

My second question concerns your typology of models. Would you include models of the past, such as the Garden of Eden, which cannot be realized again? Where do you work this in?

Wartofsky: These are both good questions. The first one I will begin to answer, although I haven't thought it through fully: I don't like to sound apocalyptic about this kind of social experimentation. I believe that one should never *force* anyone to experiment, but one should permit any experiment that anyone wants to make. In this sense, one can say that the serious responsibility for undertaking the experiment is inherent in the precariousness of present human experience, in the depth of its present ills. This obliges us to undertake to shape a future different from the present.

Mazlish: At what level are you talking? We could experiment with different ways of living, and with different types of families. Furthermore, what happens when you get into the large-scale social processes?

Ozbekhan: Technology is beginning to permit us to make these experiments off-line; that is, through various types of simulation, social experiments are becoming more and more feasible.

Mazlish: This is proliferation of experiment.

Wartofsky: Yes, with no risks.

Tillinghast: Did you say that any experiment ought to be done?

Ozbekhan: As long as it is off-line and doesn't hurt anyone.

Wartofsky: That, of course, is a big qualification. Furthermore, unless we had a way of knowing that any experiment we allowed would end up with a *less* optimal solution, in the long run, than if we did not allow it, then our deciding not to allow it is a presumption.

Tillinghast: But the Nazi doctors tried a number of experiments.

Wartofsky: I wouldn't call these experiments. The doctors in the extermination camps did not do experiments. There was nothing they were trying to discover about the future. These were perversions of the experimental process. It is a real and a serious problem, and the question of costs in human lives of an experiment going wrong or haywire is very difficult. I think this is why Kant argued so hard for a criterion of morality independent of consequences of actions. He said that since we can't possibly know what all the consequences of any act are, our decision that an act is good can't depend on what its consequences are but has to depend on something else, namely, on whether it is in accordance with some imperative that any rational person could will to be obeyed by every other person.

In answer to your second question, I don't think that the Garden of Eden myth is the re-creation of a Paradise lost, of what once was and can't be again. Whatever the historical or anthropological sources of such myths may be, I think they are essentially interpretations of the future. I can't see them as meaningful in any other sense, except perhaps with respect to notions of lost innocence. It is interesting that in all the versions of the future, as in the versions of the lost past or of the Golden Age, you end up with a kind of unity between nature and man, and man and man, which is supposed to have existed ideally before Eve tempted Adam. Any sense in which such myths are meaningful as models, or as normative or ethical or psychological insights, interprets them as visions of the future. There may be a much more profound interpretation of them, of course. But I think this much is true of them. They are not simply symbolic representations of a historical past, or records of racial memory. There is one other sense in which I think this is interesting. Many of our models are models of the past, so they would not, at first glance, seem to be envisioning the future. But I think that, even here, conceiving or reconstructing the past in a certain way is already the expression of a belief in how things ought to be or ought not to be. Thus, the historical reconstruction of the characteristics of a past age – that is, a purportedly descriptive account of the past as far as we can reconstruct it – is normative, in the sense that any historical reconstruction is selective. Certain things are regarded as important by the person who does the reconstruction. So, in a sense, even such historical models mirror present interests. But it is precisely such present interests and their vectorial character that enter into models of the future.

Jungk: I would like to ask you once more: Are there any places where such models are being built? I mentioned earlier in my presentation of 'Mankind 2000' that, in Germany, I started a series of books devoted to the 'building' of societal models. I would like to have other conferences and books grow out of them and to cooperate with similar enterprises in other countries, but I have not yet found any place where new models or new visions are actually being produced. If we really think seriously about model building, we might make a distinction between pure social imagination and applied social imagination. The first would concern itself with very 'far out' visions – not something that has to be done right away. You have pure model building, pure visionary construction, so that you might also have practical model building derived from actual, and pressing, necessities.

The visionary or the prophet, who was a very important person in former ages, will somehow have to be brought back into our society and into our context of things; and he must not be regarded as a fool or as a useless dreamer. We certainly should have institutions where models of desirable futures can be built; then we should not only construct models answering pressing needs but go beyond that and try to embody in a future-oriented society the visionary tradition from which philosophy, the arts, and most of our creativity spring.

Wartofsky: I agree, but aren't we missing the fact that we have an enormous source of that kind of things available to us already which we are not utilizing fully? If we can say that any model, whenever it is constructed, is a model of the future if it carries this kind of heuristic function with it, then we have an untapped source of such models in any number of areas of thought. Certainly this is true in philosophy. Thus, even Aristotle's and Plato's works contain models for the future (not simply in their time but now) in the sense that there is more in them than they could have foreseen. In this sense, these are creative models, as I described them in my paper.

I have argued in another context that metaphysics is a heuristic for science, a source of models for science. I will also argue that in many contexts, there are things that we do not take to be models which function in this way, heuristically.

We spoke earlier of revolutionary political parties. There is much revolutionary activity in the world today which represents this kind of model building, but perhaps not in the specific theoretical sense of having formulated an ideology or plan, or anything of this sort. I think

students in the university are creating models right and left, that is, experimenting with modes of action, proliferating them in a way that hasn't been characteristic of university life for a long time. In this extended sense of 'model' there is this kind of model building going on all about us.

The need here is for conscious realization, utilization, and a readiness to face the risks of it. Now, the explicit articulation of these things in some systematic way is necessary. One doesn't just sit down to think up models, just as one doesn't simply *decide* to become a prophet. Furthermore, if there were a prophet who was not thought to be crazy or a dreamer by a large section of present society, then he would actually be a very ineffectual prophet, because his message wouldn't be at all revolutionary. He would be a 'social worker' in the worst sense, and certainly not a prophet. (I am not against social workers; I am talking about them in a metaphorical way.)

Ozbekhan: Several of the points you have made are now finding an operational description, especially in the strategic and operating planning area, by something I have called 'causative anticipation' – the calculated anticipations that are causative of action. These are decision modes that a real plan, a futures-creative plan, ought to contain.

Wartofsky: I would like very much to see the material on that. There is a very subtle consideration of this (in *The Critique of Pure Reason*, of all places) in Kant's notion of one of the types of causality – the relation of community, or so-called reciprocal or functional causality.

Jungk: I think you are right when you say that the prophet has to go against the grain of society, but the tremendous difficulty of the prophet nowadays is that he is not taken seriously by the intellectual community. In former ages there existed a number of people who were ready to see the prophet as a saint, even if they did not understand what he said. Today a prophet is almost automatically classified as a charlatan by his community. I wonder if this does not betray a certain attitude of the academic community that is directed against the radically new, against the unusual, against the future. Hasn't the elite become too narrow-minded, too fact-bound to accept or even welcome the prophet?

Wartofsky: In the case you describe, I would say he is not the right prophet for his time. I mean that he is not effective. He would have to be effective within *this* context, able to transform this present. He can't cry 'foul' on the grounds that if he had existed in another age, he would have had a following. A prophet, by definition is a prophet in his own time, and not simply a forecaster.

From the Floor: Mr. Wartofsky, I don't understand what you meant when you spoke of this conference as a model for the future. Would you clarify that for me?

Wartofsky: It is a model in the special sense I described, namely a model of the kind of thing one does when one wants to talk about inventing the future. The model itself becomes an instrument for *creating* the future environment in the very process as you begin organizing and talking about this, whether successfully or not. That is what I think I had in mind in the subtitle of my paper, 'Models as Modes of Action.'

NOTE

[1] It is striking that biography is perhaps one of the most revolutionary of the arts in this respect. How many world leaders in history have been molded, in their aims and modes of action (for better or worse), by biographical works such as Plutarch's *Lives* ... ? This is an essential feature, too, of the creative models that history and literature present. A superficial and simplistic version of this view is to consider such works simply as didactic. This reduces the creative model to a trivial parable.

FROM PRAXIS TO LOGOS: GENETIC EPISTEMOLOGY AND PHYSICS[1]

[1971]

I. GENETIC EPISTEMOLOGY AND PHYSICAL SCIENCE

Let me begin with a claim: Genetic epistemology is exactly what it claims to be. It is *not* developmental psychology. Genetic epistemology is a theory (or the domain of theories) about the growth of knowledge. Its theoretical claim is that developmental psychology is relevant to the study of the growth of knowledge, and in this respect, it is a radical theory. It is a standard view in the history and philosophy of science that the psychology of concept formation, or of theory formation, in the sciences is to be sharply distinguished from the epistemological analysis of theory and theory change; that, whereas the psychological study of the genesis of concepts and theories may give us insight into the mind or the behavior of scientists, it is ultimately irrelevant to the objective question of the epistemological status of such concepts and theories as knowledge or truth claims about the physical world. At one time, the distinction was sharply made between 'contexts of discovery' and 'contexts of justification,' and epistemology proper was seen to be concerned only with the latter. (Here lies one source of the divison between 'philosophy' and 'psychology,' which Beilin talks about.) But the claim of genetic epistemology is radical on precisely this count; it asserts that the growth of knowledge may be studied in the psychology of concept and theory formation, and moreover, that scientific though has its genetic origins in the typical structures and stages of cognitive growth in childhood.

Let me make a stronger claim: for Genetic epistemology, there is no distinction in principle between the history of physics and developmental psychology. Rather there is a division of labor between those who study the growth of knowledge phylogenetically, in the context of the history of the species (that is, its *cognitive* history, as against, say, its biological or social or economic history), and those who study the ontogenesis of cognitive growth in typical individuals. Piaget's enterprise, seen in this

light, is only incidentally 'psychology,' in the normal acceptance of this term (if there is a 'normal' acceptance). It is rather part of a much bolder theoretical enterprise, more akin to Giambattista Vico's historicism, or to Hegel's *Phänomenologie des Geistes*, or to Marx's historical materialism, or to Spencer's evolutionism, than it is to the small-bore experimentalism of 'child psychology,' in any of its typical disciplinary cubbyholes. Piaget's treatment of the development of physical concepts acquires its intellectual force and dramatic quality from his overriding developmental theory. The drama comes from a powerful tension between a neo-Kantian emphasis on species structures, as universal genetic patterns or forms of cognition, recapitulated by each individual of the species in each generation, and a neo-Hegelian emphasis on the unfolding (literally, 'de-velopment') or maturation of species knowledge in the history of science. The tension is there despite the fact that the latter is a tacit framework, whereas the former is the explicit context of Piaget's experimental work. But the *telos* of Piaget's developmental psychology is scientific truth; the child's stages are, recursively viewed, steps towards a 'correct' world view of physical reality, such as is achieved in adult scientific thought. This places Piaget squarely in the tradition of developmental historians of science of the French school (Meyerson, Lalande, and Brunschvicg), as well as in the traditon of that Mach for whom scientific thought was a biologically adaptive instrument of human action, and who saw the history of physics as a systematic development of the methodological principle of the 'economy of thought.'

Let me repeat the claim made here in another way: *The history of the development of physical concepts and theories is the history of physics; and genetic epistemology deals with the recurrent and underlying human conditions for the development of physics.* Is the claim, then, that the child's conception of objects, of causality, of space and time, of number, is part of the history of physics; that the developmental psychology of concept formation is to be woven into the 'pre-history,' or the 'prescientific period,' of the development of physical thought?

Perhaps the difficulty lies in the understanding of the term 'history.' If 'history' is understood as a reconstruction of the past of the human species, and if 'physics' is understood as systematic theory construction, observation, and experiment, dealing with matter (its constitution, structure, laws of motion, and change), then the developmental

psychology of physical concept formation is an ontogenetic concern, with phylogenetic overtones only in the thinnest sense: namely, that the individual is an ideal type of the species and therefore simply exemplifies species characteristics. That is, such an ontogenetically oriented developmental psychology has to do with the species as species – with its biological and psychological structures and functions, once they are already evolved. *History*, on the other hand, presumably begins where speciation ends. Biological evolution having achieved a certain adaptive structure, ends (or at least our concern with it ends). To the extent that ontogeny recapitulates phylogeny, the individual simply realizes its species character in its development. On such a view, it would seem proper to distinguish *history* from *genesis*; and the history of physics from the genesis of physical concepts in the individual, or in the species. The latter is therefore a *species* study (akin to anatomy and physiology), and is, in a certain sense, part of biology – albeit of a ramified biology which deals with cognitive structure and function of the species. (The legitimation of such an approach in biology is already contained in the characterization of the species as *Homo Sapiens*.) The genesis of physical concepts, on such a view, is not an historical study. The history of physical concept and theory formation presupposes the species already formed, the species structures of cognition already at hand, and simply goes on to recount or reconstruct the temporal adventures of the species in developing physical theory. Thus, the history of physics and the developmental psychology (or biology) of cognitive growth are seen as utterly distinct.

   Beilin begins his account by reporting that Piaget prefers to be characterized as an epistemologist rather than as a psychologist. Let me take this seriously, as Beilin intends, and then argue that if it is true, the foregoing dichotomy between *history* and *genesis* will not hold up. If, as Beilin points out, Piaget's claim to be an epistemologist rests on his view that the nature of knowledge is related to how knowledge is acquired, then the history of physics and the genesis of physical concepts in the child are part of one and the same enterprise. But this is an exceptionally strong claim: namely, that *history* is *phylogeny*; that stages of historical development of physical theory are, in effect, stages of phylogenetic knowledge acquisition by the species, and in this sense are to be counted as part of its natural history, or at least, as an extension of its natural history, as a cognitive species. But then, the curious result is that *ontogeny recapitulates history*, that the ontogenesis of physical concepts

and theories recapitulates the cognitive history of the species acquisition of physical knowledge. More specifically, the claim is that the ontogenesis of physical concepts is a part, and the recurrent and underlying part, of the history of physics; that it is the common starting point, the eternal womb of physical theory; and that the very condition for understanding physics and for its continued growth is the common species character of its fundamental concepts of space, time, and object, and the concomitant concepts of number and causality.[2]

This is a heady thesis, both for its speculative breadth and for its integrative potentialities. It is not simply that the child develops into the man, but that the child develops into Man, that is, into a species individual; and that an understanding of the history of science bears on (and is borne upon by) an understanding of concept development in the child.

The talk about ontogenesis and phylogenesis masks a more fundamental point: What is arrived at, both by the developing individual, and by the species, in its cognitive growth is knowledge of the external world. Therefore, what is involved here is not simply the genesis of concepts taken as a descriptive, empirical study. Rather, the study is normative, and teleological. For the claim is that the development of physical concepts is a progressive and adaptive one – that it eventuates in *knowledge*; that the conceptual network and the theories thus evolved approximate more adequately to truth. Thus, the study is not simply one which concerns the history or genesis of physical concepts, but their adequacy as well. This is the force of the epistemological emphasis in Piaget, and in this sense, he is the heir and executor of the tradition which includes the historian-epistemologists of science like Mach, Meyerson, Lalande, and others. But Piaget turns the enterprise to distinctive considerations: How is the genesis of physical concepts, as an epistemological concern, to be studied not alone in the framework of the history of physics, but in the living recreation of its foundations in every generation – that is, in the child's achievement of a conception of the physical world? In short, Piaget's genetic epistemology asks for an account of the mechanisms, structures, and functions of human thought which make physics possible, but distinctively asks for an account of their origins and early development in childhood.

In a sense, Piaget is asking Kant's question in the First Critique: "How is Science Possible?" But where Kant's epistemology gives a transcendental deduction of what structures of mind would be necessary

for science to be possible, Piaget's genetic epistemology gives a developmental account of such structures.

The developmental account, however, differs widely from Kant's in one crucial respect. Though it is bound to something like conditional or biologically evolved *a priori* conditions for knowledge and for concept formation, Piaget's developmental account takes its departure from the phenomenology of *praxis*, from the concrete circumstances and experiences of action. Like Kant, Piaget knows where he has to get to; the articulated concepts of *space, time object, causality, number*. The forms of perception, the concepts of the understanding, stand for Piaget, as for Kant, as the *telos* of the theory. There is must attain. But Piaget's methodology is, ultimately, not Kant's transcendental deduction, legislated by what a rigorous and sustained rationality finds necessary, *a priori*, for the possibility of science – that is, what reason reconstructs as the necessary conditions for rational comprehension of the physical world. Piaget's methodology is closer to Aristotle's functionalism, asking rather: What would an organism, so constructed and in such a world, require for its preservation and for its growth? What instrumentalities of action, habit, thought, would equip it to fulfill its 'nature'? – its 'nature' being nothing more mysterious than its typical patterns of maturation, its well-being, its sociality (vague but plausibly definable characteristics of the species). Piaget's notion of 'stages' is anything but a simple, empirically derived description of the move from action to concept. Rather, it is a teleologically conceived theory of maturation, of typical growth of the individual, from neonate organism to species being, that is, to cognitive adult member of a scientific community. The *telos* is quite clear: it is *truth*. Its mechanism is equally clear – it is what Piaget calls 'true search,' but might just as well have called the search for truth, since what characterizes 'true search' is its methodology, and the norm for this methodology is one or another version of scientific, or prescientific inquiry. Without this *telos*, one could reduce Piaget's genetic epistemology to descriptive developmental psychology. But *knowledge* is a normative term; its acquisition is not simply a natural process (though it is certainly also that), but an achievement. The stages of cognitive growth are thereby not simply sequential, but developmental, cumulative, and adapatively successful. Cognitive development is not simply a *fact* of ontogenesis, but a *norm* of ontogenesis. But the norm is not simply ontogenetic, not simply set by the biologically a *priori* structures of the organism, as biological

individual. Rather, the norm is set by the cognitive development of the species, in the history of science. A Newton, a Planck, an Einstein simply represent what the capacities of the species are, in attaining knowledge of the physical world. The account of the child's development of physical concepts is an account of the early and typical development of these capacities.

That is how I understand Piaget's enterprise. These prefatory remarks therefore set the context for the discussion of Beilin's thoughtful and well-elaborated reconstruction and critique of Piaget's views on the development of physical concepts. At the same time, these remarks address themselves right off to Beilin's conclusion that "it is not clear that the data of psychological research . . . have been of any significant use to philosophers" (*p*. 000). I have already suggested what the relation is between philosophy and developmental psychology – or more accurately, developmental psychology in the context of genetic epistemology. Put this way, Beilin's question about the testability of the claim that understanding the nature of mind "is the same" as understanding how it develops (*p*. 000) almost answers itself. Developmental psychology is the applied, experimental inquiry into the conditions and typical structures and sequences of cognitive growth; it is, in effect, applied or experimental epistemology, as distinct from the related scientific enterprise, analytic epistemology. Its theoretical context is genetic – that is, it is epistemology viewed in the contexts of the origins of knowledge, and of the evolution or growth of knowledge; further, it is essentially a naturalistic epistemology with a strong instrumentalist bias, in that the attainment of knowledge is seen as an adaptive utility in normal maturation of the individual, and in the history of cognitive development of the species. The origins of our concepts and theories in action, in the *praxis* necessary for human life, and the redirection of our concepts and theories to the rationalization of *praxis*, both attest to the functional character of theory in the life of the species. In such a context, philosophy and developmental psychology can hardly be seen as exclusive, but are rather different emphases within the same enterprise.

This is a roughly reconstructed Piagetism, and it may strike psychologists as odd, even if they are aware of Piaget's own insistence on the term 'epistemology,' and his philosophical biases. (It often strikes competent and hard-working physicists as odd that such otherwise sober scientists as Einstein, Poincaré, or Bohr should have permitted themselves to waste so much useful professional time on matters

'philosophical,' when they might have been doing *real* work.) But without Piaget's notions of *genesis* and *structure*, without the framework of epistemology proper (as against 'psychology' – a distinction which Piaget makes and violates constantly), his notion of *physical concept* cannot be realized.

A physical concept is one whose appropriate or fully developed instance is to be found in physics, that is, in the most current physical science. The evolution of that concept is to be traced in the history of physical thought, that is, in the history of physics proper. And the origins of that concept are to be found, presumably, in the prehistory of physics, that is, in the history of primitive thought, at some 'beginning' which arises out of human action, or practice, and out of the human capacities or structures with which such practice is undertaken. But since we have little access to prehistoric man, we study this genesis analogically in children, *who may be said to recapitulate the stages of this prehistoric conceptual development*. This is, in effect, what Piaget (1970) calls "the fundamental hypothesis of genetic epistemology."

Now what is normative and what is not, in the study of the growth of knowledge? One may say that any study of 'growth' is already imbued with a teleological spirit, as least in that qualified sense of teleology which views growth as a normal pattern, with an ordered sequence of stages, in which the earlier stages are preconditions for the later ones. Such a 'law of development' is not teleological in the sense that it is foreordained with respect to some conscious end in view; nor does it require conscious goal-seeking at any stage. The organism need not be said to 'know' its end in view, any more than one would ascribe such a knowledge, say, to the foetus in its stages of development and growth. But such a pattern of growth is ordered with respect to a 'normal' sequence, and as such, it is lawlike. This sort of developmental pattern, or law of development, has been characterized as 'teleonomic' rather than 'teleological,' by Pittendrigh, to distinguish such developmental patterns in biology from conscious goal-seeking with an end in view; and it seems clear that this is the kind of 'teleology' which Aristotle has in mind in his naturalistic and nonmental account of organic development (see Wartofsky, 1968, p. 261). Piaget may be said to have a similar notion of growth, normative in this qualified sense, when he distinguishes 'psychological' from 'epistemological' questions. That is, apart from the normative question of *truth*, or of *truth seeking*, one may deal with normal cognitive sequence as an 'empirical' question. One need

not, therefore, ascribe greater approximation to truth in the later stages, or, to put it differently, the pattern of cognitive growth may not be, and need not be, interpreted as a pattern of epistemological growth. The latter would require a criterion of growth of knowledge such that one could determine a passage from ignorance to knowledge. In short, the latter would require a definition of truth and a criterion of truth, and the ordering of stages of growth would be determined by this criterion, rather than by any temporal or maturational sequence. Now the two alternative sequences may be identified, or they may be confused. A psychological sequence of perceptual and conceptual maturation may be taken to be identical with, or at least isomorphic with, a sequence of epistemological attainment, on some explicit theory of their correlation. But it is more likely that the two will simply be confused one with the other. Piaget tries very hard to keep them apart, and therefore to distinguish genetic psychology from genetic epistemology. He separates the question of epistemological achievement from the question of psychological development, by being explicit about the criterion of conceptual success; he takes it to be, in effect, the successful socialization of the child with respect to the norms of the adult community. In cognitive terms, this means the conceptual socialization of the child with regard to the norms of adult common sense and adult natural science, mathematics, and logic. But he leaves the epistemological status of these adult norms open. For example, in *The child's conception of physical causality* (1960), he poses the question of the relation of child thought to reality (or external reality) as a 'psychological' rather than an 'epistemological' problem, on the grounds that he is adopting as a norm the view of world reality of contemporary science and contemporary common sense, but is not intruding upon the epistemological question of the status of this scientific world view. Piaget writes, "For our part, we shall confine ourselves to psychology, to the search, that is, for the relation between child-thought and reality as the scientific thought of our time conceives it. And this view, narrow and question-begging though it appear, will enable us to formulate very clearly several outstanding problems" (1960, p. 238). But he adds, shortly thereafter, the following, on the relation which this 'restricted method' may have to a wider theoretical program:

It may very well be that the psychological laws arrived at by our restricted method can be extended into epistemological laws arrived at by analysis of the history of the sciences: the elimination of realism, of substantialism, of dynamism, the growth of relativism, etc.,

all these are evolutionary laws which appear to be common both to the development of
the child and to that of scientific thought.

   We are in no way suggesting, it need hardly be said, that our psychological results
will admit straight away of being generalized into epistemological laws. All that we
expect is that, with the cooperation of methods more powerful than our own (historical,
sociological methods, etc.), it will be possible to establish between our conclusions and
those of epistemological analysis a relation of particular case to general law, or rather
of infinitesimal variation to the whole of a curve (Piaget, 1960, p. 240).

However one takes Piaget's methodological qualifications, or his
separation of psychological from epistemological considerations, the
wider theoretical program sets the framework for the narrower
methodology nevertheless. Piaget eschews dogmatism in avoiding a
simplistic claim for the 'truth' of the adult world view, or of
contemporary science. Yet, it *is* his norm, and it is his norm on genetic
grounds. Science is the outcome of species adaptation; it is, so to speak,
an organic development, deriving from the interplay of *praxis* and the
structures which the human organism brings to its experience. Therefore,
Piaget has put the epistomological question into a developmental
framework as well. What criteria of rationality, of criticism, of analysis,
of practical test *can* there be for epistemic claims, which are not
themselves the products of human action and reflection? If the
epistemological status of the contemporary scientific world view is itself
to come under critical scrutiny, it can only be by a cognitive apparatus
evolved under the very conditions studied in genetic psychology. And
insofar as such a reflective conceptual criticism is part of the enterprise of
the history of science, of historical sociology, of sociology of knowledge,
of analytic epistemology – that is, of critical reflective disciplines whose
subject matter *is* this very domain of adult thought itself – to that extent
too is Piaget's *own* study a contribution to epistemology; and where this
is explicit, it is the content of his genetic epistemology. In fact, Piaget's
explicit hope is that the psychological studies of child thought may
themselves contribute to the clarification of *scientific* questions. Thus, he
recounts the effect of Einstein's question to him on the primitivity of
time or speed, in the child's conception of the relative motions of objects:
Einstein's question, in 1928, set off a whole series of psychological
studies by Piaget and his collaborators; these psychological findings were
later utilized in the work of the French relativistic physicists, Abel and
Malvaux (1954). Piaget uses this example to make a point concerning the
relation of child psychology and genetic epistemology:

In certain instances, the genetic study of the construction of concepts and operations provides a response to questions posed by the sciences with respect to their methods of knowledge. When this is the case, child psychology becomes extended into a 'genetic epistemology' (Piaget, 1967, pp. 83–84).

Strictly speaking, one may question this 'extension,' on the grounds that even here, the wider epistemological question as to the validity of the scientific world view, or its 'truth,' is not touched on, but only the correlation of child thought with the articulated and systematic thought of the adult world as expressed in science. One may argue that the contribution which genetic *psychology* makes to science is simply one of suggesting fresh conceptual standpoints, derived from the child's thought, as interesting starting points for theoretical construction in physics. In this sense, the reconstruction of the child's world view, in a systematic fashion, is analogous to the interpretation of classical metaphysical constructs in the history of philosophy, as themselves a heuristic for scientific thought and theory construction (see Wartofsky, 1967). But I would argue in defense of the *epistemological* content of such reconstructions and interpretations – they are part of the ongoing analysis and critique of concepts, which is linked to the formulation of testable hypotheses, and this *is* an epistemological enterprise which concerns epistemological issues in the philosophy of science.

In this sense, the relation of *Praxis*, of concrete practice and action in the world, to *Logos*, to the construction of conceptual systems or theories which serve both as rational explanations and as guides to action, is similar in the genesis of the child's thought and in the genesis of scientific thought. Piaget has already suggested analogies between child thought and, for example, the genesis of the physical world view of the pre-Socratics (Thales, Anaximander, Anaximenes, and others; see Piaget, 1931). But these thinkers were the generators and founders of physical science, and not simply 'childish philosophers.' If an antigenetic view would argue that one should separate the genesis and history of science from science 'proper' and from the epistemological questions in science, it would also argue that genetic psychology and genetic epistemology are separate. The same questions arise in both cases, indicating that the general issue is the same, and that the relation of genetic psychology to genetic epistemology bears on the relation of the history of science to the epistemology of science, or more broadly, to the philosophy of science (an issue which has finally reared its head among

philosophers of science). I would go further and suggest that Piaget's model of the reciprocity between genesis and structure, and his account of the activity[3] of the subject in constructing such logical, mathematical, and physical structures, is as viable for the 'genetic epistemology' (or historical epistemology) of science, as it is for the genetic epistemology of child thought.

   To sum up: The relation of genetic epistemology to the development of physical concepts needs to be seen not simply in the context of the development of the child's conception of physical causality, or of number, or of space and time, but in the context of contemporary physical science, mathematics, and logic, *and* in the context of the history of science. At the minimum, this is required because the stages of cognitive development are seen by Piaget within the framework of socialization of the child – that is, of the achievement by the child of the world view of adult common sense and contemporary science. (Where these latter two are in conflict, more interesting problems and opportunities arise.) But beyond this, if Piaget's work is seen as having epistemological import, then the development of child thought becomes an aspect of the development of species knowledge, and as such it is part of either the prehistory of physics, or of the ongoing cognitive ground for the cultural continuity of physical science; that is, the child's development of such concepts becomes the cultural or social 'apprenticeship' without which the cultural transmission of scientific thought, and the provision of intellectual cadres for the continuation of scientific thought, could not take place. This may be seen as having only descriptive import, as a matter of the sociology of science or of knowledge; but in fact, the attainment of knowledge, as a quest for truths about the world, is the content of scientific inquiry. And insofar as the child, in its cognitive development, is made a participant in this quest, the epistemological import of its development is clear, though the epistemological questions concerning this development still remain to be studied. Piaget liberates analytic epistemology from the narrower constraints of formalism, and from the dogmas of analytic self-sufficiency, by bridging the gap between child psychology and genetic epistemology. That the bridge remains problematic attests to the difficulty of the question, but not to its irrelevance.

II. ON PIAGET'S 'MENTALISM'

There is a fast and loose characterization of 'mentalism' abroad in the

land, both in philosophical and psychological discussion. Beilin is correct in talking about a "renaissance of mentalist philosophizing in psychology," if we accept his very broad characterization of 'mentalism' – namely, the ascription of 'cognitive mechanisms' to the organism, or the introduction of 'terms of great generality,' like *equilibration*, or of more limited applicability, like *conservation*. On these grounds, anything beyond functional correlation of data statements is 'mentalist,' and even the hardiest behaviorist is tainted with mentalist heresy, once he strays from measurement statements and their correlation. The confusion here seems to be between 'mentalist' and 'inferred-entity theorist.' The latter is a mentalist only if the 'inferred entities' (mechanisms, operations, capacities, etc.) are 'mentalistic.' But this depends on some specified demarcation of 'mental' from 'physical' (or, at the very least, 'nonmental'), and the issue remains unjoined if the demarcation is not specified. For example, *drive* is not a 'mental' term, but an *abstract* term, presumably a functional term, characterizing a disposition, or a set of dispositions. But unless 'disposition' in turn is characterized as 'mental,' there is no ground for characterizing a drive theorist as a mentalist.

There is a more restricted sense of 'mental' which Beilin adduces, however, which derives from 'inferred entity' talk, but which specifically avers 'covert' or 'internal' cognitive operation, and these, presumably, are operations of the mind. Now there is a certain liberation from the self-conscious constraints of philosophical behaviorism, in freely talking, once again, of the 'mental.' But the strength of the claim which such usage makes, or the characterization of the operations or structures which it suggests, has to be carefully assessed. This has to do with general problems of theory construction or model building. If Piaget talks about *equilibration, conservation, reversibility, assimilation*, or *accommodation,* these are clearly theoretical terms of wide import and of great abstraction. But they are just as clearly functional terms, that is, general terms for specifiable operations, introduced both to order and to explain a large variety of particular operations. All instances of a certain operation can be characterized as instances of reversibility, whatever the specific contexts, because of formal similarities in the operation. So too with the other terms. What tempts us to characterize these terms as mentalistic is that these operations are ostensibly covert, internalized, and take place 'in the mind' at a certain stage of development, though at an earlier stage, as action concepts, they are exemplified in outward

behavior, in the manipulation of objects, in body, hand, or eye movements. So too with the notion of *structures* of the mind: functions presumably bespeak structures, as the abiding or developed forms which account for the functions. Structures of the mind are reducible therefore to dispositions, faculties, or capacities, evoked appropriately in interaction with an environment, and attested to by the regularity or universality of their evocation under similar environmental conditions. But what makes such structures 'mental'? In a trivial sense, they are 'mental' just because that is what we mean by 'mental' and nothing more. But then, 'mentalism' is a fake issue, since even the most physicalistic antimentalist will find no great discomfort in talking about 'structures' as typical values of a set of response variables under typical stimulus conditions. If what makes such general terms for operations or structures 'mentalistic' is something more than this, we have to know *what* more. If it is the speculative or constructive latitude which Piaget permits himself in his introduction of such general and abstract terms, then 'mentalism' is no more than an index of theoretical imaginativeness. The *S* matrix in quantum mechanics is, on this account, just as 'mental' as equilibration in Piaget's genetic epistemology. If the objection is that, in the one case, we are talking about physical entities *tout court*, and in the other, about cognitive ones, then the distinction is simply a handy one, and 'mentalist' comes to no more than 'cognitivist.' So what? That Piaget is a cognitivist is no great discovery. And to permit the perversities of an old-fashioned and ill-advised behaviorist reductionism to dictate that all who refuse to be held within its injunctions are hereafter to be labeled 'mentalist' is simply to substitute one perversity for another. Perhaps "mentalist" should be used thus: physicist, biologist, mentalist – that is, as a synonym for psychologist. Since even Skinner eschews the 'empty organism' view, what is left?

So much for Piaget's 'mentalism.' It ascribes to him only the methodological freedom to construct *theories*, rather than to simply 'report the facts.' But since this has been the prerogative of every other science, and moreover, is a condition for its even claiming to *be* a science, what is all the fuss and bother about in psychology? I would venture to guess that the issue *does* concern theory construction, about which psychologists have been self-conscious and shamefaced for two generations, simply because they have had to reject *some* classical theories as less than fruitful, if not downright misleading. So what? Piaget's theoretical constructs may prove to be false, misleading,

fruitless over the long run. But he is a scientist and so he theorizes. He is a psychologist, and so the theoretical constructs he proposes are 'mental.' And so he is a 'mentalist.' But then, so is every psychological theorist, in this vacuous sense of 'mental.' Nor do we need the Fodorian apologies, that, in fact, there may be neurophysiological structures to 'explain' the mental structures or operations, so that one may be a 'mentalist' and a 'materialist' all at once. Since when is it the case that methodological diversity entails metaphysical dualism? Now, that there are deep issues here, I will not deny. But they are not the ones raised in the contemporary hoopla over 'mentalism.' Chomsky introduces 'mind' by methodological fiat, as a theoretical framework for the theory of language acquisition (and also as a rhetorical ploy within an uptight discipline, in order to *épater les bourgeois*). Piaget proposes theoretical terms for 'mental' operations, because he is dealing with knowledge and its growth, and 'knowing' may harmlessly be described as 'mental,' in a grand old tradition. The fetishism of 'mentality,' which Ryle takes as the object of his special critique, and which has become a local industry (small manufacturing) in Anglo-American philosophy, should not blind us to the vacuousness of the charge of 'mentalism' in Piaget's case. Being "squarely among the mentalists in that these mechanisms and processes are inferred entities created to explain observed behavior" (Beilin, p. 000) is no more than being 'squarely among the scientists who create theories to explain the way the world is.'

III. FROM 'MENTALISM' TO 'INNATISM'

More crucial than confusions made or distinctions drawn concerning 'mentalism' is the real issue involved in *nativism* or *innatism*. There is a weak sense of *innatism*, which surely characterizes Piaget's view. If the terms *assimilation* and *accommodation* are to make any sense at all, then there must be some structure or character to *which* the child assimilates or accommodates his new experience. The logic of the process requires, therefore, a structure prior to experience, to which it is assimilated or accommodated. But this is a weak sense indeed, since one can argue that such structures are not native but are the acquisition of earliest experience; in effect, the imprinting, on some almost totally plastic organism by its original, neonate experience. Now this obviously goes too far, for we are not born genetically innocent, but are indeed the heirs to our species evolution. The question, therefore, is how to

interpret claims about *cognitive* structures. Perceptual mechanisms and structures, according to Piaget, have their roots in organic processes, but are developed in interaction with an external environment. Cognitive structures are developed in this interaction, not apart from perception, but as an aspect of the development of perception. As Beilin makes clear, Piaget clearly distinguishes his own position, on the genesis and status of such cognitive structures, from that of the empiricists (associationists or inductivists, like G. E. Müller) and from innatists or apriorists (even among the Gestaltists – Piaget specifically mentions W. Metzger – who disclaimed his characterization). Piaget writes, "it is necessary to oppose the geneticism without structure of empiricism and the structuralism without genesis of Gestalt phenomenology with a genetic structuralism in which each structure is the product of a genesis and each genesis merely the passage from a less evolved structure to a more complex one. It is in this context of an active structuring that the exchange between subject and object take place" (Piaget, 1969, p. 364).

Beilin correctly criticizes the views which ascribe innatism to Piaget. But he puts the burden of *apriorism* on Piaget's notion of fixed developmental stages, and characterizes him therefore as a 'preformationist' or a 'maturationist.' If there is, in fact, a 'fixed' and 'built-in' developmental sequence, "impervious to environmental influence" with respect to its sequence structure, though facilitated by transactions with an environment, this is innatism enough. Indeed, Beilin ascribes such a sequence (if it were to exist) to a species-specific genetic programming; such a notion of a genetic program impervious to environmental influence (except perhaps in the facilitation of these built-in stages) is as innatist as one can reasonably get, without going all the way, to a notion of 'innate ideas' which includes specific 'content of idea' as well as 'form.' But Beilin dismisses, as untenable, the alternative possibility of "consistent cultural experience in a wide variety of social environments." Here, perhaps, Piaget needs defense. For, after correctly characterizing Piaget's view as neither innatist nor empiricist, and after correctly assessing Piaget's emphasis on the activity of the subject, and on interactionism, Beilin still ascribes a strong *apriorism* and innatism to Piaget, by way of his maturationism. But Piaget is quite explicit in rejecting such a view, which he characterizes as "structuralism without genesis." He sees the strongest influence of such a view in Gestalt psychology, and he sees it as stemming from the work of Husserl (specifically, the influence of Husserl, 1913). Piaget describes Koffka as

holding such a view, namely, that "development is determined entirely by maturation, i.e. by a preformation which itself obeys Gestalt laws. Genesis remains secondary to the fundamental preformist perspective" (Piaget, 1967, p. 146). In contrast, Piaget characterizes his own view as decidedly different in that it requires "true reciprocity of structure and genesis," and therefore no preformation. His thesis is that "genesis emanates from a structure and culminates in a structure." Thus, there is no fully plastic organism, as a radical associationist-empiricist might claim; at every stage there is *some* structure, but a structure plastic enough to develop, that is, to be modified or reconstructed in the interaction with an environment. Conversely, therefore, "every structure has a genesis," says Piaget; and, "there are no innate structures: every structure presupposes a construction." This seems clear and explicit enough as a rejection of innatism in the sense of preformationism. But the question may still be asked, as Beilin asks it: granting that there are no preformed structures of perception or of cognition, is there a preformed maturational sequence which is itself a *second-order* structure that is innate? Or does Piaget's anti-innatism hold also for the maturational sequence? Can the sequence be *changed* – that is, not merely slowed down, speeded up, or repressed? Can there be jumps in the sequence? Are there alternative sequences? In the criticism of Koffka, cited above, Piaget seems straightforward in his rejection of a fixed and preformed maturational sequence. Yet it seems clear that the scientific aim of Piaget's work is to establish certain laws of maturational development, certain 'fixed' sequences of perceptual and cognitive attainment.

Here the issues and confusions on the question of innatism all rise to the surface (as they have in the discussion of Chomsky's 'innatism' in linguistics), and it is worth sorting out just what the claim (or the charge) of innatism contains. The strongest claim for innatism is that our 'ideas' are innate, or that 'language' is innate. But of course, the strength of the claim depends entirely on what one means by 'idea' or 'language.' The strongest innatism might claim (as in Plato's case) that the 'idea' of *circle* is innate, as a precondition for the experience of something as a circle; and that such 'ideas' are universals of which the particular experiences are instances or exemplars. In the linguistic cases, the strongest claim would be that *a* language is innate, that is, some natural language, say English or Chinese. But this is patently absurd. The modified innatist claim is that universal formal structures of language are innate (as

'species-specific genetic programs,' perhaps?), in the sense that, as Chomsky might put it, grammatical competence is a necessary condition for language acquisiton, and that such grammatical competence bespeaks a grammar, as the universal formal structure which is mapped onto the surface structure of speech (thus, mapped differently onto various phonetic, syntactic, and semantic 'surfaces' in different natural languages, but, at the level of its deep structure, universal for all languages). There is a reading of this view that makes it a plausible hypothesis: namely, that there is a biological component of the evolved organism which may be characterized as structural, and which is genetically determined, or at least, genetically cued or specified. Thus, it is biologically *a priori*, and in this sense, innate. Thus far, Piaget would agree, I think. But what exactly is such a species-specific structure? It is plain enough,if we take various organs of the body as our examples. The lungs, for example, have the 'competence' to effect a transfer of oxygen and carbon dioxide simply as a function of their physiological chemical structure (that is, in conjunction with the heart, circulatory system, the structure of blood cells, the characteristics of gas diffussion through a membrane, etc.). But there does not seem to be a *development* of lung function in maturation (except in the fetal stage, and in the transition to air breathing at birth), though there is growth. However, in language acquisition, as in perceptual and conceptual growth, there is a pattern of development which may be characterized as a sequence of stages, in which elements of the later stage are not present in the earlier ones. What is 'innate' in such a case cannot be the acquired language, or perceptual skill, or the acquired concepts. We are forced back, therefore, either to (*a*) an underlying formal structure which perdures through such acquisition, and provides the permanent framework through all the developmental stages (for example, universal grammar, in its 'deep structure' or Kantian *a priori* forms of perception or of the understanding); or (*b*) an underlying pattern of the development of structures, that is, a second-order structure which remains permanent or constant (or fixed) as the *form* of first-order structural change, that is, as the specification of the sequence stages of structural change. Such a second-order structure is, in effect, a law of evolution or of development. But what sort of 'structure' is this? It is clearly not to be understood as an organ, or as a physiological structure in the ordinary sense. To call it 'genetic preprogramming' makes use of a metaphor whose import is not clear in the case of cognitive or conceptual growth, though it does make

sense to talk of genetic control or specification of a sequence of maturation in biological terms – for example, stages of myelination of neurons, of the development of the musculature of focus and accommodation in the eye, or of hormonal regulation of physical growth patterns or of sexual maturation. These latter may be characterized as 'preformed' in terms of 'genetic preprogramming.' But does Piaget allege that this sort of developmental sequence is characteristic of perceptual or cognitive growth? Is preprogramming or preformation requisite for all developmental or evolutionary laws? Apparently not, for he specifically inveighs against such a view (for example, in Koffka, as we have seen, but also in his rejection of Kantian *apriorism* in its transcendental, or 'necessary' form). Piaget holds, rather, that the reciprocity of structure and genesis has to be a 'genuine' reciprocity, that is, such that the genesis actually generates new structures by modification of older ones. On such an interactionist view, the only thing that can 'fix' a developmental sequence is the joint operation of structure and environment, in the crucible of practical and critical activity by a subject. Since genesis is a necessary condition for the emergence of structure, and since genesis and prior structure (or equivalently, structure and subsequent genesis) are the necessary and jointly sufficient conditions for any new structure, as Piaget urges that they are, then 'imperviousness to environmental influence' cannot make sense, and we are driven to seek our universals of developmental sequence elsewhere than in 'genetic preprogramming.'

Piaget holds a more dialectical and less mechanistic view that the one Beilin ascribes to him, I think. The fixity of developmental sequence does not require preformation, but does require universals of *both* structure and environment; and both of these may plausibly be held to be provided in any normal human context. What such a fixed developmental sequence may be said to be impervious to are *accidents* of environmental influence – by definition, therefore, those environmental variations do not, or cannot affect the sequence, in its ordinal structure, but can only cause variations *within* the maturational order – that is, speed it up, slow it down, repress some elements of it, elicit others, etc.

I see a difficulty with this formulation too, however. It introduces, in effect, a *caeteris paribus* clause into the account of development, which simply underwrites 'normal' developmental patterns (that is, those which have a high distribution in the experimenter's sample, or worse yet, those which he chooses to take as 'fair sample' or paradigm cases) as the *universal* cases, or patterns of development.

There is still another conceptual difficulty with the notion of maturational sequence as a universal and 'fixed' sequence of stages, if the fixity is taken as a product of the joint operation of structural and genetic universals: Given some sequence, *ABCDF*, which is reconstructed as the 'normal' case, and given some alternative sequence *ABDEF*, in which the element *C* does not appear, and in which a new element *E* appears, do we then say that we have two 'incomplete' mappings of some prefigured structure, *ABCDEF,* one of which is 'normal' simply on grounds of distribution? Or do we have two alternative sequences, with only certain similarities, and no prefigured or preformed 'deep structure' of which they may be said to be 'incomplete' mappings? Which of the two sequences represents a law of development? Piaget, I suspect, would answer with Leibniz, *Natura non facit saltum*, and that 'skipped' stages simply do not occur, because (in our example) *D* in the second sequence just could not be the same stage as *D* in the first sequence, since it was not generated out of the same previous structure. (The *D* which is a modification of *C* is not the same as the *D* which is a modification of *B*.) We may compensate, in experience, for missing structures, as, for example, the blind do. But the result is a different structure from that acquired by the sighted. Still, where the compensation has to do with *cognitive* structures, there may be equivalences, so that the same set of formal properties of a group of operations (for example, in logical or mathematical concepts, or in the basic formal concepts of reversibility) may be achieved. But the physical intuitions in the modality (for example, sight) will simply be missing, and will be substituted for, to the extent that there is an isomorphism at the perceptual level, by physical intuitions (and operations) in another sense modality (for example touch).

In short, it seems to me that Piaget need not be characterized as a preformationist, even in the qualified way which Beilin suggests, but should be given his due as an interactionist. The 'fixed' maturational sequence requires fixity in both the native structures which the organism brings to its experience, and in the structures of that experience itself, as they are features of the objective world of the subject's activity: of his physical and his cultural and human world. Socialization (conceptul socialization included here) is not merely a matter of imposing some *a priori* sociality of man's innate nature upon his experience; nor is it merely a matter of fully plastic adaptation to the demands of an objectively sundered world which imprints its 'own' patterns on us. And

it is too much to assume that the pattern of socialization – for that is what cognitive growth is, in its deepest implications – is genetically preprogrammed. But, it therefore seems to me too much to assume that Piaget assumes it, for the logic of his argument provides for formation without preformation, for sequence and law without design, and for the development of concepts without innate determination of their form or content or sequence. What remains, as the most general thesis concerning the emergence of physical concepts, is a meta-theory of development, which guides Piaget's work, and which is so easily accepted and so easily acceptable to common sense that its audacity and its heuristic power are too often overlooked: namely, the thesis that *praxis* and *logos* are indissoluble, and that in the genetic sequence of human development, the separation of *logos* from *praxis* is impossible.

NOTES

[1] Discussion of an article by H. Beilin, 'The Development of Physical Concepts,' in T. Mischel, ed., *Cognitive Development and Epistemology*, New York and London: Academic Press, 1971, pp. 85–119.

[2] This would presumably hold as well for other contexts. The ontogeny of social development or of socialization would be seen to recapitulate the phylogeny (the species history) of socialization or social formation, of property right, of communality, of rules and laws, etc.

[3] On the activity of the subject as 'practical activity,' Piaget notes that the spirit of Kant's proposal that perception is 'organized' by the subject should be taken not in a 'transcendental' way, as Kant took it, but in terms of the 'real constructions' which arise in actual perceptual activity. He sees in this the content of Marx's *Theses on Feuerbach* (see, especially, Thesis V), in which "practical, human-sensuous activity" is counterposed to "sensuous contemplation" (Piaget, 1969, p. 362, footnote 1). By extension, given Piaget's account of the relation of perception to conceptual thought, this same 'practical activity' is the matrix and genesis of conceptual-theoretical structures, as these are, in turn, the organizing modes of practical activity, in that reciprocity of genesis and structure which Piaget sees as his distinctive view. The relation of this view to Hegelian and Marxist notions of dialectic interaction of subject and object seems *prima facie* clear, but needs to be examined further.

BIBLIOGRAPHY

E. Husserl, 'Ideen zu einer reinen Phänomenologie und phänomenologische Philosophie.' In *Jahrbuch für Philosophie und phänomenologische Forschung*, 1 (Niemeyer, Halle; 1913).

J. Piaget, 'Children's philosophies.' In C. Murchison (Ed.), *Handbook of child psychology*. (Clark Univ. Press, Worcester, Massachusetts, 1931).

J. Piaget, *The child's conception of physical causality*. (Littlefield, Paterson, New Jersey; 1960) (reprint).

J. Piaget, *Six psychological studies*. Edited by D. Elkind. (Random House–Vintage, New York; 1967).

J. Piaget, *The mechanisms of perception*. Translated by G. N. Seagrim. (Routledge, London; 1969).

J. Piaget, *Genetic epistemology*. Translated by E. Duckworth. (Columbia Univ. Press, New York; 1970).

M. Wartofsky, 'Metaphysics as heuristic for science'. In R. S. Cohen & M. Wartofsky (Eds.), *Boston studies in the philosophy of science*. **3.** (Humanities Press, New York; 1967). pp. 123–172.

M. Wartofsky, *Conceptual foundations of scientific thought*. (Macmillan, New York; 1968).

PICTURES, REPRESENTATION, AND THE UNDERSTANDING

[1972]

I am tempted to subtitle this paper 'Meditations on a Dinner Plate', partly out of respect for Gombrich's essay, 'Meditations on a Hobby-Horse'; partly out of a disenchantment with the British preoccupation with pennies and their elliptical looks; and partly out of the conviction that a philosopher is likely to do better on a good meal, than on a pocketful of small change.

It should be clear from all this that I shall be concerned in this paper with a famous (or infamous) philosophical shape: the circle, which allegedly appears elliptical when viewed obliquely. I propose to examine and to uphold the thesis that the tilted circle, which is supposed to appear elliptical, in fact *appears* circular; but that we represent it as elliptical in our pictures in order to *make it* appear circular; and therefore, that talk about elliptical appearances is talk about how we represent what we see, and how our seeing is affected by our canons of representation. In short, we picture what we understand, when we represent titled circles as ellipses in our pictures. So we may be said to see what we see by way of our picturing, when we claim to see a tilted circle as an elipse. All this is packed rather densely here, and I hope to unpack it so that it becomes clear, and so that any sense of paradox or perversity will vanish.

The philosophical proponents of elliptical looks are the sense-datum theorists, from Moore, Broad and Russell, through Price, Ayer, and their elliptoid American counterparts. What I want to claim is that the sense-datum theorists, in talking about elliptical looks, have literally taken leave of their senses, and have constructed for us what is in fact a philosophical counterfeit – a fiction of the understanding – which has been passed off on an unsuspecting populace as true sensory coin. In rejecting, or at least reinterpreting their claims, I hope to generalize from the discussion of ellipses and circles to a thesis concerning the part played by representations and pictures in our knowledge of the external world.

Now, beyond these opening remarks, some brief background: The argument from illusion, on which much of sense-datum theory builds, runs something like this: When we look at a plate or a coin from any

angle off the normal, what we know to be a round or circular object appears elliptical. Since the elliptical appearance cannot in fact be part of the surface of the object, and since such appearances change as we move about, while the object or its so-called 'real shape' is supposed to remain the same, what appears to us cannot be the real shape of the object, but rather a sense-datum, or a series of such sense-data. In Price's elaborate construction, we are introduced to families of such sense data, some of which are so-called nuclear sense data, constituting standard solids and some of which are parts of distortion series of this family. Russell's version talks of a *space of perspectives,* in a frankly geometrical construction, which relies very little on sense-data as appearances, but relies rather on the laws of geometrical optics and of perspective transformation. Early Ayer, of *The Foundations of Empirical Knowledge*, reduces the sense-datum theory to a 'manner of speaking' in which there is allegedly no difference in the facts which sense-datum talk and material-object talk are both about, thus evading an ontology of sense-data. But all are taken with elliptical looks as somehow given or known by acquaintance.

Now let us ask a very unphilosophical question: What are the facts of the perception of tilted circles? How could one tell what a tilted circle appears as? Plainly, we all know what it means for it to appear as an ellipse, and all of us would presumably exhibit our visual sophistication by *representing* or *picturing* such tilted circles as ellipses. Thus, it has been very easy to become the victims of the naiveté of our sophistication, and to concur that, after all, what is given in perception (as anyone can plainly see) is an elliptical appearance. We can, after all, 'see' the ellipse, even though we 'know' it is a tilted circle. Thus, the classical view held that it was some perceptual judgment which preserved shape constancy, 'interpreting' or 'correcting' the elliptical appearance. Helmholtz, for example, spoke of this as an *unbewusstes schluss*, an unconscious inference or judgment. But how is one to tell *what* appears, from which such an inference, conscious or unconscious, is made? Just looking and seeing apparently will not do, if the inference is unconscious, or if it necessarily intrudes on the looking, for then one can never get an 'uncorrected' or 'uninterpreted' appearance. And unless one has an independent mode of access to the appearance, one is in fact making a hypothetical inference *to* the appearance, rather than making an inference *from* it; that is to say, one is performing what Peirce described as an abductive inference to 'elliptical looks' as a premise from which it

would follow (on the additional premise that the perceptual judgment that preserves shape constancy is an inference) that the real shape is circular. But if there is no independent mode of access to the appearance, or to what is 'given' as a sense-datum, then (1) it cannot be 'given' in any perceptual sense, but only in the judgmental sense of a hypothetical premise, and (2) there is no logical reason why the premise should be that the 'given' shape is elliptical, rather than square, or triangular, or pear-shaped. Each of these, through some set of transformations, will yield a circle: and any of them will do as the hypothesis from which, by the 'proper' *ad hoc* transformation, we arrive at the 'corrected' judgment that what we see is a tilted circle.

There obviously is some reason why we choose the elliptical shape as that which 'appears'; or rather, there is a set of reasons. The set of transformations that we choose to employ is that of geometrical optics, i.e. that set of projective transformations which, when the projection is made onto a plane surface, give us the so-called laws of perspective in drawing. Such a projection would yield an ellipse from a tilted circle. Furthermore, since vision is a matter of reflected light from surfaces (which is presumably projected through an aperture with an inverting lens onto a plane surface, i.e. the retina), retinal image-formation is taken to be such a projection onto a plane surface. Our neurophysiological model of vision and of the eye, adopts this geometrical optics, therefore. This is reinforced by the fact that, were we to look at the image formed on a retina (say in an eye removed from an animal, like a rabbit), the *image* of a tilted circle would in fact be an ellipse.

But there are serious objections to such a simple view of image formation, even at the level of the neurophysiology of vision. The fact that the retinal projection is on a curved rather than a flat surface; that the so-called 'surface' is really an array of highly specific neurons, which fire differentially; that so-called visual projection areas on the retina are complex rather than simple mappings of the projected light rays (i.e. that contour, edge, and slant receptors for 'shape' are not simply arranged); and finally, that we do not 'see' our retinal images with some third or inner eye – all these are well-known complications of the simple picture theory of visual perception. Thus, one could claim against the thesis that we 'see' what is projected onto our retinas: first, that we just do not; and second, that this account of seeing is theory-laden, dependent on our beliefs and our knowledge of geometrical optics and the physiology of vision. Thus, seeing tilted circles as ellipses is seeing through theory-tinted glasses.

I think this is true, but it is not the whole truth. For the account of what appears to us when we look at tilted circles may be given by someone who is not theoretically committed to all this, or who knows nothing about it. Presumably, we may find a naive subject who will tell us exactly what he thinks he sees, in a relatively theory-free way.

There are several ways to elicit this response experimentally: The experimenter may ask the subject to draw what he sees, i.e. to make a picture of it. He may present the subject with a series of shapes, from among which the subject is to pick the shape that most closely corresponds to the perceived shape. He may ask for a verbal description. (This last would presumably be the most ambiguous response with regard to differences in visual shape, although differences in shape may be expressed verbally by a naive subject, uncommitted to the theory of geometrical optics, by such expressions as 'x is rounder than y' or 'y is flatter than x and rounder than z'. But more on this later.)

From the results of a well-known set of experiments, Thouless concluded that what is seen is neither a circle nor the projected ellipse, but something in between the two. Thouless writes:

It is commonly stated in textbooks of Psychology that when we observe figures inclined to us, we see them not in the shapes indicated by the laws of perspective but in the shapes which these figures 'really' possess. Thus when we look obliquely at a circular object we see it not as an ellipse but as a true circle. While it is undoubtedly true that such an object seen in these conditions is judged to be of its true shape and also that we are prepared for motor reaction to a circular object, I do not find that experiment confirms this statement as to what shape is seen. If a subject is shown an inclined circle and is asked to select from a number of figures the one which represents the shape seen by him he chooses without hesitation an ellipse. This ellipse, however, is widely different from the one which represents the shape of the inclined circle indicated by the laws of perspective, being much nearer to the circular form. The subject sees an inclined figure neither in its 'real' shape nor in the shape which is its perspective projection but as a compromise between the two.[1]

Thus, according to Thouless, the 'phenomenal shape' is not the ellipse, as the laws of perspective transformation would represent it, but a 'regressed' ellipse, tending towards the 'real' (circular) shape. The perceptual shape constancy, which leads us to identify the tilted circle as a tilted circle, rather than as an ellipse presented in the fronto-parallel plane, thus intrudes, according to Thouless, upon the presented stimulus-object, which Thouless identifies with the ellipse 'correctly' presented as a perspective transformation according to the laws of geometrical optics.[2]

Thouless's point, of course, is that so-called shape constancy is *not* preserved, but that there is instead what he calls 'compromise' between the *stimulus-shape*, i.e. the perspectival projection on the retina, and the *real shape*, for which our motor responses are cued. Thus, the 'compromise' *phenomenal shape* is a 'phenomenal regression to the real object'. By contrast, when Thouless's subjects were asked to reproduce or to match actually elliptical disks or trapezoidal figures, there was practically no such 'distortion' of the stimulus shape; that is, the ratio of long to short axes of the ellipses drawn or matched were very close to those presented.

Among the many experiments on this problem reported since then, it has been shown that as the depth-cues (e.g. light or texture gradients) or surround-cues were removed, shape constancy decreased. In effect, what such experiments do is impoverish the visual presentation of the tilted circle so that the actual visual information in the input approaches that of a presented ellipse in the frontoparallel plane. Eissler[3] removed depth cues by monocular presentation. Langdon[4] showed that shape-constancy is inversely proportional to removal of perceptual cues, such as surround-cues. Leibowitz and Bourne[5] showed that shape-constancy varied directly with exposure duration and luminance – i.e. as visual acuity was impaired, shape constancy decreased.

Constancy was also shown to vary with intelligence and training. For example, Thouless reported that experienced artists and teachers of perspective, though they show some 'phenomenal regression', match the stimulus-shape more closely. That is to say, they preserve constancy least, in their reproduction or matching of the stimulus shape. Likewise, Leibowitz, Waskow, Loeffler, and Glaser, in their study[6] ('Intelligence Level as a Variable in the Perception of Shape') showed that constancy effects were strongest for Rhesus monkeys, then decreased in order of 'intelligence', among the other groups tested (mental defectives, slow learners, first-year psychology students, and Ford scholars making up their rank-ordered groups of subjects).

In general, then, the claim of these experiments is that a tilted circle 'appears' as an ellipse to the degree that visual acuity or normal visual cues are impoverished, and to the degree that 'intelligence' or training intrudes upon the stimulus-input. Further, the experiments tacitly or explicitly presume to be getting at what the subject 'sees' by contrast to what he is presented with, as an independently definable stimulus-object, or stimulus-shape. Thouless's distinction, between 'real shape',

'stimulus-shape', and 'phenomenal shape' re-articulates the standard difference, in studies of perceptual constancies, between 'real' and 'apparent' shape, but introduces the mediation of stimulus-shape. The problem here is, of course, the independent specification of the stimulus, apart from the 'phenomenal' response. Here, as in psychophysical experiments generally, the stimulus is defined within the context of some physical theory. Just as in psycho-physics, stimulus-strength is measured within some such theoretical framework as electrical energy, or luminance, or temperature, here the stimulus-shape, as projective ellipse, is 'measured', or defined by the laws of geometrical optics. The 'real' shape, to which the constancy is relevant, on the other hand, is defined by its tactile-motor properties, i.e. in the three-dimensional, Euclidean-Cartesian space taken to be canonical for physical objects. The plane-projections of geometrical optics are transformations of this Euclidean space, which are embodied in the laws of perspective. Further, Cartesian optics and dioptrics construct a physiological model of the eye according to these same laws. It is this model that defines the stimulus-shape, in this case, as a projective ellipse.

   Now the question arises as to what defines the so-called phenomenal shape, or the 'perceived' shape, which the experiments show to vary from the ellipse? It is obviously not enough to say, as Thouless does, that there is a 'compromise', between the real shape which perceptual constancy ought to yield and the 'phenomenal' shape. For this states the problem entirely in terms of visual stimulus-variables and some presumably tactile-motor response-variables, implying that the 'compromise' is simply one between sense-modalities. But Thouless's own results, with teachers of perspective, and those of Leibowitz *et al*, suggest strongly that what intrudes here is visual *understanding*. What I shall mean here by visual understanding is a familiarity with and a knowledge of canons of visual representation. Such an understanding is presumably acquired – it can be taught, and it can be learned. More than this, it can be induced in the experimental situation itself, by appropriate verbal cues. My argument, in short, will be that the choice of the elliptical shape as the 'appearance' of a tilted circle is a function of a commonly understood canon of representation, in drawing, namely that of linear perspective. Now it would be vacuous to claim that tilted circles are drawn or represented by artists as ellipses *because* they have adopted the laws of perspective as canons of 'correct' representation. This verges on tautology. (What it means to have adopted such canons, is to represent

tilted circles as ellipses.) But if we are concerned with experimental subjects who are not artists, and who are asked to choose matching shapes, rather than to draw them, the question arises as to how a canon of representation in drawing operates on such presumably naive subjects. The experimental situation is worth examining here.

In all of the studies cited, the subject is asked to reproduce or represent what he sees by choosing a matching shape from among a set of such shapes. (These may be cut-out shapes, or pictured shapes). Thouless writes, for example, that the subject is asked to choose a shape that 'represents the shape seen by him'. Thus, the very condition for getting access to what 'appears' to the subject is bound up with (1) *the subject's representation of it*, and with (2) *what he is asked to represent by the experimenter*. Thus two things are involved here, both of which conspire to produce a representation according in some degree with the canonical representation of the tilted circle as an ellipse, and therefore in violation of the preservation of shape constancy. The 'regression', on my view, is not *to* the 'real object', but *away* from it, and to a canon of representation.

The argument for this is as follows: (1) First, with respect to the subject's representation of the perceived shape, the very condition of representing the shape involves what one takes a representation to be. The naive view is that a representation somehow reproduces the perceived shape of the object. The logic of the experiment (and of the term 'phenomenal shape') assumes this, since the subject's representation is explicitly taken to be the 'perceived shape'. But I would argue instead that a representation does not reproduce the 'perceived shape. But I would argue instead that a representation does not reproduce the 'perceived shape' (whatever that may be) but rather specifies a canonical similarity (from among an infinite set of such similarities). That is to say, the representation is a conventionally adopted specification, which looks 'right', or is a 'proper' representation, by virtue of our acceptance of a certain 'vocabulary of forms' (Gombrich's phrase, I think). In this case, the appropriate convention is the elliptical 'picture'. Why then do the untrained subjects produce a 'regressed' ellipse tending to the circle? There is a lot buried in this sort of question: (1) First, there is the residual assumption that what makes the ellipse the appropriate convention is some 'natural' tie it has with the stimulus-image. But this, I have argued earlier, will not hold up, because we do not 'see' the stimulus-image, if we mean by it the retinal-image; but also, the 'picture' we have of a

retinal-image is itself a 'picture', though in this case a neurophysiological picture, or model, derived from seventeenth-century optics and dioptrics. Second, there is in the question the assumption that the shape-constancy which exerts influence on the visual image is a function of tactile-motor perception, and is not native to the eye, so to speak. But whatever the intermodal effects, I think this does not explain the subject's 'regression to the real object'. Rather, I would argue that the 'compromise' Thouless speaks of is between what the subject sees (a tilted circle) and what he knows he should represent it as (an ellipse); and that the compromise is a function of skill in the use of the canons of perspective representation, i.e. in visual understanding. The argument gets complicated here by the fact that the subjects are not asked to *draw* the shape presented to them, but rather to choose a representation of it. But I would argue that these two are closely related. That is, the visual-manual skill of *drawing* the ellipse in accordance with the laws of perspective is directly related to *seeing* it in terms of such a representation. The experimental results bear this out in the following way: teachers of perspective chose shapes closest to the so-called 'stimulus-shape', in Thouless's experiment. Monkeys and mental defectives chose shapes closest to the 'real shape', in the experiments of Leibowitz *et al.* In neither case was the drawing skill called upon. What was called upon was the visual understanding of a canon of representation. But this understanding is in direct relation to the practice and the acquired skill of drawing in accordance with the canon. The normal subject's 'index of regression' is therefore an inverse measure of this skill, on this view.[7] The subject's representation is therefore, in general, a function of his visual understanding, defined in this case by his practical knowledge of the laws of perspective; this 'knowledge' defined, further, as his ability to draw shapes in accordance with this canon; and directly related to this, his ability to choose from among a set of shapes the ones that are in closest accord with this canon. Therefore, the subject may be said to 'see' by way of his picturing.

(2) Second, however, what the subject chooses to represent is also a function of what he is asked to represent by the experimenter. Here, the analogy between impoverishment of the visual presentation by removal of perceptual cues, and impoverishment of the visual presentation by verbal cues, is significant. In the experiments by Eissler, Langdon, Leibowitz, and Bourne cited above, luminance, duration of exposure, surround cues, etc., were controlled in such a way that the stimulus-input of the tilted circle approached more and more closely that of an ellipse. But a similar effect can be induced by verbal directions to the subject,

deliberately, or inadvertently. If the subject is asked to represent the
'shape' that is presented to him, or to match the 'shape' he sees by
choosing from among a set of two-dimensional, abstracted 'shapes', cut
out or drawn, then what is being elicited is already an abstraction from
the visual field. The question 'what is the shape that you see' or the
direction 'choose a matching shape' already impoverishes the visual
field, by directing attention to an abstracted part of it, just as decreased
luminance or exposure time do. If I name the object, for example, and
ask the subject simply to 'draw the dinner plate you see before you', I
would venture the guess that a less elliptical figure would be drawn, in
most instances, than if I asked the subject to 'draw the shape of the top
of the dinner plate'. This latter question impoverishes the perceptual
field, and directs the subject to remove surround cues 'subjectively' or
'intentionally', so to speak, rather than removing them objectively, as in
the other experiments. The artist, trained visually to fix upon the
abstracted property 'shape', and trained in perspective drawing, will be
able to approach the so-called 'stimulus-shape' in his representation,
precisely because the abstracted property he is fixing on is defined by the
same laws of geometrical optics that define the 'stimulus-shape'. But this
is part of the history of the development of the laws of perspective, about
which more needs to be said than can properly be said here. Suffice it to
note that 'shape', apart from the 'real object' (of which it is only one
property, and a very complex one at that), is already an abstraction,
selected as much by the history and development of our representational
and picturing activity, as by what is presented to us in the visual field.

    The untrained child, drawing a tilted circle, preserves constancy, on
the other hand. The child's drawing of a plate on a table looks something
like this: (Figure 1).
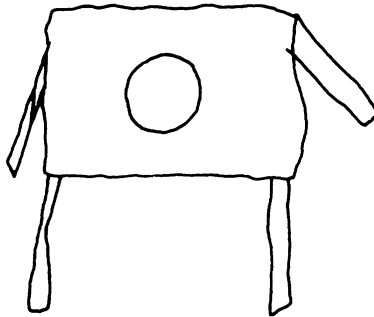
Fig. 1

Neither the trapezoidal foreshortening of the rectangular table, nor the elliptical foreshortening of the circular plate, enters into the representation. But note what happens when the 'proper' representation of the table in perspective is combined with the shape-constancy preserving circle (Figure 2). We would 'read' this picture as that of a plate standing upright on its edge on the table. The surround cues of the table, its grain, its shortened rear legs, all work against the visual understanding of the plate as lying *on* the table. So too, in Figure 3 we see the ellipse as a tilted
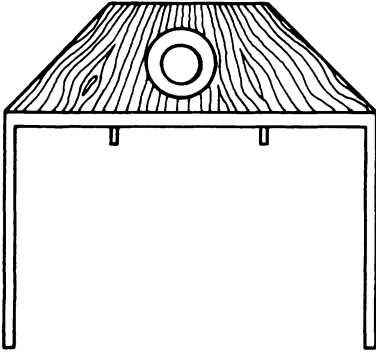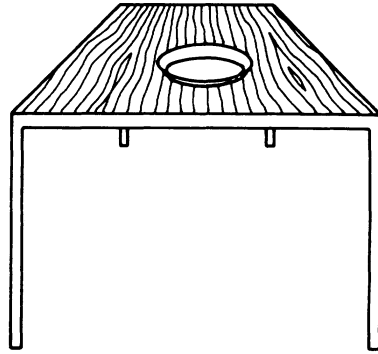
Fig. 2   Fig. 3

circle, giving the perspectival surround cues of the table, etc., whereas if we were given simply the abstracted shapes of a rectangle and an ellipse (Figure 4), no such interpretation would be given. Figure 5 *suggests* an interpretation like that of Figure 3 but less forcefully so, since 'table'
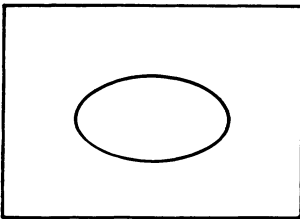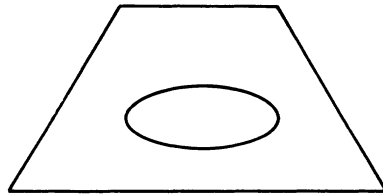
Fig. 4   Fig. 5

cues, (legs, grain), and 'plate' cues (thickness of edge, etc.) have been removed. Through all this, what is working most powerfully is the

convention of representation which maps three-dimensional objects onto a plane surface. This is a very complex transformation. But we take it without difficulty, because the canon is so well established. (So too, we accept unflinchingly the radically divergent canons of representation which give us Popeye, Mickey Mouse, and the huge-headed children in the comic strips 'Peanuts' or 'Miss Peach'.)

The secret of our acceptance in these exceedingly odd cases is familiarity, not simply with this or that character as drawn, but with a style, or a style-complex. But in the case of the 'correct' drawing of the projective elipse, our familiarity is somehow normative in a different way. Or so it would appear. No one would argue that any human being could be constructed as Popeye is, or as Charlie Brown is. Yet, caricature yields resemblance, and even insight and unique depiction. But in the case of 'correct' perspective representation, we take the canon to be verdical, and not simply an accepted convention. On my argument, the two cases are similar. Yet, I think, everyone would feel that the perspective representation is actually 'closer to the way things really look', or further still, that it *is* the way things look. This is, if not an epistemological fact, certainly a cultural fact, and it needs coping with. But to cope with it properly, I would have to go into the status of the canons of perspective drawing more fully than I have in the suggestions made thus far. This is the burden of a continuation of this essay, still in work.

To summarize thus far, however: I argue that in normal perception of objects, and in the normal *praxis* which is the context of this perception, shape-constancy is the correct account of what we see. We see tilted circles as round, i.e. as tilted *circles*, and not as ellipses. Only by abstraction do we separate the shape from the object, so that it can be taken as itself an object of perception. But what intrudes on our doing this is the condition of communicating what we see, in this abstractive way, by means of a representation of it. Once we are required to *represent a shape*, we are already caught up in the web of canons of representation – and in particular, in the modern world, we are bound to a canonical framework derived from geometrical optics, both with respect to our models of the physiology of vision, and with respect to our representations of what we see, by way of pictures. The same geometrical optics that gives us the ellipse as a retinal image of the tilted circle gives us the laws of perspective which direct us to represent the tilted circle as an ellipse.

But if this canon is so pervasive, then in our 'acceptable' mode of representation, we 'read' the elliptical representation *as* a tilted circle, thus effecting a re-transformation in our seeing of pictures, which is the inverse of the transformation we effect in making the pictures. That is why we draw circles as ellipses *in order to make them look circular;* and that is why, if we were to draw tilted circles as circular, they would look wrong to us.

This bespeaks an important fact: that a theory as historically recent as geometrical optics is, or as perspective drawing is, has become an integral part of our visual understanding, or of our visual 'common sense', so much so that it affects our seeing. In a subsequent study, I propose to show how this took place; and also to show that we are able to entertain alternative and radically different canons of representation at the same time.

NOTES

[1] Robert H. Thouless, 'Phenomenal Regression to the Real Object', *The British Journal of Psychology*, **21**, 4, p. 339 (1931).

[2] Thouless compares this result to Hering's experiment with light intensities. In Hering's experiment, a white disk in strong shadow is seen as brighter than an intensely illuminated grey disk, despite the fact that the retinal image intensity is greater for the illuminated grey disk. So too, in Hering's experiments with constancies of hue, a blue surface illuminated by an unsaturated yellow light, so that it looks colorless to monocular vision (through a tube), continues to look blue when seen in a surround, and binocularly, (though its saturation is less than when it is illuminated by white light.) Here too, in Thouless's terms, there is "regression to the 'real' object", i.e. to 'real' brightness or intensity of a surface, or to 'real' hue. (But this is more problematic, perhaps, because of the greater ambiguity of what we mean by 'real color' as against 'real shape', so that the reference-object or reference-properties of the perceptual constancy is not as clearly specified in the color (brightness, hue) case as in the case of shape.)

[3] K. Eissler, 'Die Gestaltkonstanz der Sehdinge bei Variation der Objekte und ihrer Einwirkungsweise auf den Wahrnehmenden', *Archiv für Gestaltpsychologie* **88**, pp. 487–550 (1933).

[4] J. Langdon, 'The Perception of a Changing Shape', *Quarterly Journal of Experimental Psychology* **3**, pp. 157–165 (1951).

[5] H. Leibowitz and L. E. Bourne, Jr., 'Time and Intensity as Determiners of Perceived Shape', *Journal of Experimental Psychology* **51**, 4, pp. 277–281 (1956).

[6] H. Leibowitz, I. Waskow, N. Loeffler, and F. Glaser, 'Intelligence Level as a Variable in the Perception of Shape', *Quarterly Journal of Experimental Psychology* **11**, pp. 108–112 (1959).

[7] Talk about 'skill', or 'correct' representation is taken here to be relative to a given canon – i.e., the laws of linear perspective. Cézanne's departures from this canon are

not to be taken therefore as lack of skill, but rather the deliberate choice of another canon. The same thing is true, I think, of all those art forms that 'violate' the laws of perspective, e.g., Romanesque painting, or classical Chinese or Japanese drawing and painting.

PERCEPTION, REPRESENTATION, AND THE FORMS OF ACTION: TOWARDS AN HISTORICAL EPISTEMOLOGY

[1973]

I. INTRODUCTION: CRITIQUE OF AHISTORICAL THEORIES OF PERCEPTION

Among the things which are generally taken to change, historically, are ideas, theories, social systems, technologies, customs, beliefs. Biological evolutionary changes or developments are often distinguished from and sometimes compared with these historical changes: Thus, species-change, or the evolution of particular organs or traits, or even geological change are taken to be processes of *natural* transformation, as distinct from those *post-natural* or *cultural* changes which may be characterized as historical, and which involve human action and human history distinctively. Thus, for example, cultural evolution is contrasted with natural or biological evolution, the 'noösphere' is contrasted with the 'biosphere', and transhistorical species-specific traints, such as erect posture, or speech are contrasted with culturally variant features, such as particular natural languages, or even more differentially, styles or customs or political systems.

Human perception has been studied, and conceived as a species-characteristic – in effect as a transhistorical or even more simply, a biological capacity. Its structures and modes have been understood as ahistorical, (though *what* is perceived obviously varies historically). This has been the practice, certainly, in physiological and psychological studies of perception. It has been reinforced, if not in fact even determined in some measure by the traditional philosophical analyses of perception aś a universal human faculty, and by the relation which philosophers have established between perception and knowledge, in various epistemological theories. The view, in various epistemological theories, is that though knowledge may grow and change, and evidence (and what counts as evidence) may also change, the empirical base for such knowledge, or its test in our perception of the world remains anchored in an unchanging and universal human perceptual capacity. Though there is no question that perception is a universal human faculty,

and that it is crucially related to epistemological contexts, I will argue in this paper that the forms or modes of perception, its structures themselves, are historically variant; that this variation is related to *historical* changes in the forms or modes of human action (or *praxis*): and that this variation or change is perceptual modes in both determined by, and in turn helps to determine such historically changing modes of human action. Furthermore, I will argue that if this is true of perception, then several traditional philosophical characterization of epistemological questions are wrong, and that what is needed to replace them is an *historical epistemology*.

The theses I will argue for, in examining the foundations for such an historical epistemology, are: *first*, that perception itself is a highly evolved and specific mode of human action or praxis; i.e. that its characterization as *only* biological or physiological or more generally, in 'natural' contexts, is inadequate; and that moreover, its traditional treatment in philosophy, in the context of an *a*historical epistemology, is fundamentally mistaken. *Second:* that the specific feature of perception as a mode of action is that it is mediated by *representation:* and *third* that it is by the variation in modes of representation that perception itself comes to be related to historical changes in other forms of human practice, and in particular, to social and technological practice. For this argument on the role of representation in mediating perception, I want to resurrect the traditional term, *imagination*, in a specific sense, and to relate it to the activities of *picturing* and *modeling*.

At the outset, let me say how I think these theses differ sharply from those presented in other theories of perception, and also state what the problem is to which I am addressing myself.

(1) Most classical theories of perception – both rationalist and empiricist – take perception to be an ahistorical and universal species-characteristic – i.e. a general human faculty based on a common perceptual system (whether this system is taken as biologically evolved, or *a priori*, or simply taken for granted in a common sense way, e.g. as 'seeing', 'hearing', etc). I take human perception much more narrowly (or more differentially) as the specifically human faculty which develops only *after* the biological evolution of our *sensory* system has been completed. That is to say, I take it as an *historically* evolved faculty, and therefore based on the development of historical human practice. In order to sharpen the distinction I am making here, let me contrast it with three other views of perception, all of which lose this distinction:

(a) *Essentialist theories of perception*

The *general* or *abstract* philosophy of perception is mystified by a general or abstract perceptual vocabulary – e.g. 'seeing', 'hearing', etc – and conflates the generality of the terms with the generality of the activities which these terms denote. In that sense, this becomes an essentialist theory of perception, which seeks a definition of perceiving in terms of some essential relation between the activity and its objects, or in terms of some model of human perception which is historically undifferentiated.

(b) *Relativist theories of perception*

A *culturally* or *situationally variable* account of perception is given in some non-essentialist and relativist theories. Here, for example, 'seeing' is reconstrued as 'seeing *as*', and perception in general is linked to interpretation or judgment, i.e. as a (conscious or unconscious) processing of sensory input in the framework of memory, past experience, intention, cultural or situational context, etc. Thus perception is not simply an essential relation between a perceiver and perceptual objects, or an essential and unchanging structure of an *a priori* sort, but becomes a more plastic and variable activity or process of interaction, whose variability depends on acknowledged variation in context, use, background-knowledge or framework. However, this variation in perceptual mode is seen in terms of *alternative* contexts, or situations, or cultures, but not yet as a change or a development, either ontogenetically or phylogenetically, and certainly, not yet historically (though it is compatible with such accounts). One may call such views *relativist*, *pragmatic*, or *contextual* theories of perception.

(c) *Developmental or evolutionary theories*

Such theories of perception, on the other hand, do give such an account of change in the perceptual activity, or in the perceptual apparatus itself. But it is given in either genetic contexts, e.g. in developmental psychology, in which stages of perceptual change and growth are understood as general species-patterns, mapped into the stages of ontogenetic development; or they are seen, phylogenetically, as stages in the species development, i.e. in the biological evolution of the species. In short, the context, in such developmental or evolutionary theories of

perception, treat it as a species-characteristic in an exclusively biological context. Such accounts are compatible with co-variation and even causal relations between the changes in perception, (and in the perceptual apparatus) and changes in the mode of life and the practical activities of the species; and indeed, such an evolutionary epistemology sees the development of the perceptual system in the context of adaptive strategies and survival values which are selected out by natural processes. But here too, these are *pre-historical* biological contexts, and do not yet differentiate, as my theory will, between the biological and the historical contexts themselves.

To put the distinctiveness of my theory very simply, and nominally: it is neither an essentialist theory concerning ahistorical or transhistorical features of perceptual activity; nor is it a *relativist* or *pragmatic* theory of cultural or situational variability in perception; nor is it reducible to those theories which are identified with genetic epistemology or with developmental or genetic psychology; nor yet with so-called naturalistic or evolutionary epistemology. It is compatible with these latter theories insofar as they describe and explain the biological substrate of perceptual activity and its evolution, or even as they propose to account for species-wide mechanisms of perceptual activity and such species-wide features of perceptual processing which can be traced to their evolutionary contexts – e.g. perceptual constancies, cliff-effect, etc. – and which, indeed, can be shown to be common features of many species with closely related sensory-systems.

But my view goes beyond them in arguing that what the species brings to perception, as the product of its biological evolution, is the *starting point* for an historical epistemology; and that the transformation and development of this genetic inheritance is a function of changing historical *praxis*; in short, that *perception has a history*.

(2) Most *philosophical* theories of perception (with some notable exceptions) work with a model of perception based on seventeenth-century psychology. Having inherited the philosophical *problematique* from seventeenth-century philosophy – i.e. from Descartes and Locke – they remain also within the confines of that psychological model which gave rise to the characteristic philosophical problems of classical rationalism and empiricism, and have therefore never been able to transcend this particular *problematique*. In effect, such contemporary philosophical theories of perception are dealing with anomalous philosophical problems generated by an anomalous science. This is

particularly true of British and American philosophies of perception, and most specifically, it is true of the analytic school. Here the fault is especially complex. In part, it is due to the peculiar fact that a major school of contemporary scientific psychology of perception (roughly speaking, the empiricist school) is itself *also* still stuck within an anomalous seventeenth-century mechanist model of perception, characterized by a Euclidean geometric optics and an associationist logic. What I take to be anomalous here are precisely the mechanist feature of the model which confuses a particular theory of geometrical optics – i.e. a theory of the transmission, reflection and refraction of light, especially through lenses, – with a theory of vision, and in particular, with a theory of visual *perception*. However useful it has been to pursue the analogy of the eye to a camera, the reduction of the philosophical account of perception to the problems engendered by this model of sensory-physiology is surely by now anomalous; and not simply because the contemporary sensory physiology of vision has left the cruder analogy to geometrical optics far behind, but because the reduction of the perceptual model to the constraints of the physiological one confuses two distinct levels of activity. So too, the associationist logic of traditional empiricist epistemology and psychology is no longer adequate as the model even for the relations between the complex of elements involved at the level of sensory-processing; and certainly not adequate to the hierarchical organization and interaction of sensory and perceptual systems, as gestaltlike and integrative properties are discovered at lower and lower levels of the system, down to the individual neurons; and even further, to the regulatory mechanisms at the subcellular level (e.g. protein turnover and renewal, with respect to such biochemical variables as substrate concentration, nutrition, and genetic factors at the subcellular level). In short, much of contemporary philosophy of perception continues to generate its problems from an anomalous seventeenth-century model of *sensation*, in part because it is this model which now has become that of common sense. Therefore, by remaining ignorant of the relations between changing science and changing common sense, and thereby taking today's common sense to be the universal and unchanging common sense of the species, such philosophy of perception remains blissfully ignorant of its own historical limits, and the historical datedness of its models. (The evasion practised here is highly protected against correction, by the 'linguistic turn', in which the ordinary or common-sense language of perception, or its 'conceptual

grammar', is taken as a norm for our *knowledge* of perception; or more hermetically yet, when the *problem of perception*, as a *philosophical* problem, is reduced to that concerning the uses or meanings of perceptual terms or statements in ordinary or common-sense language.) This failure to recognize the historical limits of a particular theory and a model of perception is related to the ahistorical characterization of perception itself, in such ahistorical theories.

(3) This, in turn is related to the third difference between the theory I will propose here and other ahistorical theories, in particular, the traditional and contemporary empiricist or sensationist theories of perception, as well as those rationalist theories which are dualistic, like Descartes'. A whole group of theories of perception, including here sense-datum theories, representational theories, causal theories and even some gestalt theory see the 'activity' of perception as an 'inner process', in which the activity of association, or of construction, or of modes of response to perceptual stimuli are characterized as 'in the brain' or 'in the mind'; and therefore, sharply distinct from the kind of *outer* activity usually represented by our various motor-activities (including speech). Even intentionalist theories, (e.g. phenomenological theories) which emphasize the constitutive and directed character of perception and see it as actively engaged upon an *object*, nevertheless distinguish between such directedness and constitution and the directedness and constitutive nature of actual outward bodily acts. These latter alone are conceived as 'real' interventions in the natural world, whereas the domain of action of perceptual acts is taken to be that of intentional objects, themselves constituted by the act of perception, and thus not natural or spatio-temporal objects at all. This leads either to an outright mind-body dualism, or at best to a phenomenological monism, in which the natural world itself, and even its spatio-temporal features are taken to be such constructs, or constituted entities, and in which Berkeley's dictum – *esse est percipi* – hold literally. This is the idealist direction in phenomenology. The only other outcome is a phenomenalistic skepticism *cum* structures of the mind, reconstructed transcendentally in a Kantian spirit.

My own view differs sharply from all of these, in that it is an explicitly realist view of perception in two senses: *first*, that the 'objects of perception' are taken to be independent of perception, though they are mediated by the activity of perception, in that they are perceived *by means of our representations of them*. That is to say, the objects of

perception are not taken to be 'perceptual objects', but real objects – i.e.
spatio-temporal or material objects (or processes), *which* we perceive.
The mediative 'entities', which on traditional representational or causal
or sense-datum theories of perception, are taken to be the 'objects' of
perception, I take to be representations – i.e. perceptual artifacts which
we *do not perceive*, but *by means of which* we perceive real objects (or
processes). Second: by virtue of this, perception is not simply an inward
activity, directed upon some 'mental' or 'perceptual' entities 'in the
mind' or 'in the brain'; but is itself a (mediated) form of outward
activity, which is continuous with other forms of outward human action
in the world; and that even in its most interiorized modes (e.g. in
perceptual imagination, or in dreaming) it is a mode of *virtual* outward
action. Therefore, in its very genesis, perception is directly linked to that
practical interaction with an external world whose qualities and
structures are transformed by human action, and thus, by perception as
well; but which transformations are nevertheless transformation *of* an
objective and independently existing environment.

In summary then, these three features sharply distinguish the thesis I
will present from a standard group of philosophical theories of
perception: (1) I take perception to be historically variable, and not an
unchanging and universal feature of the species as such. It is universal
only in its preconditions, i.e. in terms of the biologically evolved sensory
system and the (undeveloped or native) sense modalities. (2) I reject the
seventeenth-century psychological model of perception as anomalous,
and propose an alternative model. (3) I take perception itself to be a
mode of outward action; to be derived, in its genesis, from other direct
forms of outward or motor-action or *praxis*; and to be, in perceptual
practice itself, continuous with, or a part of such outward action or
praxis. In this sense, it is perceptual activity *in* the world, and *of* a world
as it is transformed by such activity.

2. WHAT PERCEPTION IS AND WHAT IT IS NOT

Let me begin with a story: The occasion of this conference led me to
London, from which I debarked for Helsinki. At Heathrow airport, I
had the opportunity to spend some time, before departure, with two
friends who were in London. One, (a well-known anti-methodologist and

anarchist), when I told him I was to speak on perception, asked me in his characteristic fashion: "Perception of what?" Unprepared for such a question, I stumbled, "Well . . . you know, perception . . . the usual thing. In fact, I want to talk about perception and action." Little better informed, my friend and critic pursued me: "You mean that if I want to hit you, I have to be able to see you?" Now, finally, I found I could answer clearly. "No", I replied, "on the contrary. If I see you, it's *because* I want to hit you."

The moral of the story is clear and not original. The very genesis of perception is linked to its function and its uses in the life-activities of organisms. Perception is a part, and a function of those interactions between organisms and their life environment in which recognition and response to predator and prey, to danger and opportunity make the difference between life and death. Therefore the perceived world of the organism is in effect a map or an image of its activities: just as, conversely, the perceptual apparatus of various species is itself shaped to the modes of interaction by which the species survives. We do not perceive, and *then* act; perception is itself one of the instrumentalities or modes of action. That it becomes differentiated and highly specific, and that it can take place 'internally' so to speak, during a suspension of outward motor activity by the entire organism is simply one of the achieved strategies and economies of animal activity. But the relative 'internalization' of this perceptual activity, and its relative autonomy is not a systematic breach between perceptual and other forms of outward motor activity. Rather, it marks that delay between proximate causes and the organized responses of animals which permits the elaboration of information-processing systems characteristic of higher and complex organisms. In the human species, I will argue, the modes of perception, or the forms of perceptual activity are no longer bound only to the biological apparatus which has evolved in the course of species evolution. Rather, the very forms of perceptual activity are now shaped to, and also help to shape an environment created by conscious human activity itself. This environment is the world made by human *praxis* – nature transformed into artifact, and now embodying human intentions and needs in an objective way. But beyond this, perceptual activity is now also shaped to, and helps to shape a new and different 'world', namely that 'world' which is a cognitive construction, and is embodied in our representations, as theories and models in science, and as pictures in art. To say that we see by way of our picturing, or our modes of

representation, then, is to claim that perceptual activity is now mediated not only by the species-specific biologically evolved mechanisms of perception, but by the historically changing 'world' created by human practical and theoretical activity.

In short, then, I want to discuss an activist or practical theory of perception, in which perception is understood as a mode of human action; and in which it is therefore encumbered or endowed with all the qualities of human action or *praxis*, namely: effectiveness in the world, as a constituent of practical activity (causal efficacy); intentionality (as it is involved in the conscious teleology of human action); and, necessarily, a mode of physical or organic activity, involving and interacting with the physiological structures and bodily motions of the organism, and exhibiting as well the specific features of reflexiveness or internal activity characteristic of such other organic functions as digestion, emotion, or hormone balance. That this is *not* a reductive approach to perception should by now be clear. For the very foundation of what is distinctively human in perception is its character as a socially and historically achieved, and changing mode of human action; and thereby invested with a cognitive, affective and teleological character which exemplifies it as a social, and not merely a biological or neurophysiological activity. What is more, it is not an activity *of* the perceptual system or of a specific sense-modality, but an activity of the whole organism. Even at the biological level, which we share in common with other animals, it is not the organ which perceives, but the whole organism *by way of* the organ. And as a whole organism, the animal embodies not its own, or individual modes of perception, but the species-modes of perception, as they have evolved. Ontologically, of course, it is not a *species* which perceives, but an individual organism, *by means of* a species-evolved apparatus, and in a 'world' which is species-defined, in terms of the characteristic modes of activity in meeting life-needs.

Let me say then what perception is *not*, in two senses. (1) It is not the activity of an isolated or abstracted perceptual organ. (2) It is not contemplation, or passive reception of an input. I will discuss these briefly here.

(1) Were I to surgically prepare a rabbit's eye, so that all tissue functions were maintained; and were I to separate the optical and neural function of this eye from the rabbit's body, and were then to project a visual image upon the retina, so that retinal image-formation could be observed by me in the usual way, this would *not* be a case of visual

perception by the rabbit's eye. Moreover, were I to retain intact all visual activity remaining to a decorticate cat, or to an unconscious and drugged human subject, this too would not be perception. In each of these cases, we would say "the rabbit's eye does not see, nor does the decorticate cat see, nor does the drugged human subject". In the first case, *eyes* don't see; only whole organisms with visual activity see *with* their eyes; and even these do not 'see' when the visual activity does not involve the action or life activity of the organism.

(2) Let me go one step further to say what perception is not: it is not 'contemplation', on one classic model of disinterested and passive spectation; not because this latter is not perception, but because this latter, as it is described in classical spectator theory, does not exist. It is an abstractive fiction invented first by scientists, building models of vision according to the canons of geometric optics; then elaborated by philosophers (or these same scientist-philosophers) reconstructing theories of perception on the basis of these geometrical-optical models. Classic — that is to say, seventeenth-century — theories of perception converted an abstract geometry of perception into a theory of the activity itself, thereby mistaking a theory of image-formation for a theory of perception.

Now a theory of image formation is an integral part of any theory of perception, since we perceive *by way of* images. Let me make this clearer. We do not 'perceive' the images we form in the activity of perception, but rather operate or act perceptually by way of these images. We do not perceive the retinal (or aural or tactile) images formed by our sensory activity in say, vision, hearing, touch perception. But we do not perceive except by the mediation of images. I take images as representations, constructed by us, for the sake of perception; and therefore, modified by us as our perceptual activity demands; and therefore, further, as the means of instrument of perception. It is an old view, but I think still a correct one, that perceptual action, or perceptual *praxis* is a form of human (or animal) organic activity — i.e. physical activity in the life-space and life-time of an organism *mediated* by internal representation. It is therefore not a simple reflex, nor a simple effect or response to a causal external stimulus. It is a processed response, attuned to a certain end, or goal. However, the notion of 'internal representation' — i.e. of internal mapping involving selective elaboration or characterization of an external object or situation — is a dependent notion. It derives, theoretically, from the model of *external* representation, or picturing. In

the history of the theories of perception, therefore, the imaging or representational mediation is itself conceived on the historical models of then-current modes of representation. Thus, I've said that the classical (seventeenth-century) theory of perception finds its model in geometric optics and its physiological application as a theory of vision. But different theories of perception borrow from different current models of representation; and these may be mathematical; or be taken from physics; or, to an extent much greater than is usually admitted, from the forms of representation in art.

The metatheory of an historical epistemology would therefore involve the critique and analysis of *theories* of perception, and how they change in interrelation with theoretical and stylistic changes in the history of science and the history of art. But this is a separate matter. We are concerned here with how the actual forms of perception themselves change historically.

3.   THE SOCIAL AND HISTORICAL CONTEXTS OF HUMAN PERCEPTION AND THE ROLE OF REPRESENTATION

The key to the *historical* variability of perception is its link to historically variable modes of non-perceptual praxis; or more accurately, the involvement of perceptual activity *in* ordinary human action or praxis in its characteristic and historically variable modes. This may be approached in two ways: *genetically* and *reflexively*. (1) In the first case, we seek the genesis of perceptual activity itself in pre-perceptual or non-perceptual forms of action. The genesis of perception, in species-terms, or in biological-evolutionary terms would yield at most the necessary, but not yet the sufficient conditions for an account of historically-evolved perception. Thus, if we were to speak of the physiological-structural foundations of perceptual activity, e.g. in the evolution of the sensory-modalities, or in the development of the distinctively perceptual areas of the brain, or brain-function – (e.g. the visual, or auditory cortex) – we would end, at best, with a generalized, or species-specific account of the usual perceptual functions, i.e. 'seeing', 'hearing', etc., but one which is as yet abstract, or unqualified with respect to historically evolved characteristics. One might say that biological species- 'history' or species-evolution, is precisely such an account, and that perceptual development, in a phylogenetic context, yields all the 'history' of perception there is. To the extent that the

mammalian eye evolves, and to the extent that specification takes place further, from say, the higher apes to *homo sapiens*, then there is this 'history' of adaptation and selection which results in what we would then characterize as an undifferentiated or species-wide *human* perception. Further, one might argue, in terms of an evolutionary epistemology, that this adaptation itself would include the interaction of alternative perceptual strategies with changing environments. Thus, by the complex mechanisms of selection and adaptation, one may adduce at least a *prehistoric* 'history' of species-adaptation which yields the characteristic perceptual universals: e.g. the perceptual constancies (shape, distance, object, and colour-constancies, etc.), 'cliff-effect', the psycho-physical laws; and which establishes such perceptual universals in their continuity in the mammalian line, e.g. among the higher vertebrates.

But such an argument stops short just where I propose to begin: at the threshold of historical human praxis. It further assumes that the perceptual apparatus is *completed* by that time. I will argue that, though the 'apparatus' *may* be complete, it is not yet, properly speaking, the *human perceptual* apparatus, but rather only its substrate in physiological terms: what we may call the sensory-motor apparatus on which perception develops. I am therefore proposing to use the term 'human perception' in a somewhat Pickwickian way, though in a way easy to define: *human* perception begins to develop only *with* historical human praxis (to be defined shortly). Prior to that we may speak of *animal perception*, namely, that perceptual activity which we, as an evolved species, share with other animals, but which has as yet not evolved *beyond* the animal level, to the level of human culture and history. The human perceptual 'apparatus', properly-speaking, transcends the physiology of perception, or its phylogenetic development. To put it differently, the perceptual 'apparatus' includes functions which are presently at least, not accounted for in any satisfactory way by the neurophysiological account. That is to say, the *historical* development of modes of perceptual action is not yet mapped into accounts of neurophysiological structure, in present descriptions. There are suggestions that it may be, in recent research, but this is at present speculative and controversial. For example, the differential development of microstructure of the neural system under different experiential modes, (as described in, e.g., experiments with selective early visual experience of cats, early partial visual deprivation in humans, and with

orientation anisotropies in visual acuity resulting from differing visual
ecological environments); or the thesis proposed by Penfield, concerning
the effect of differing cognitive and perceptual function (focus,
attention) on the ontogenesis of brain mechanisms. So too, molecular
theories of memory suppose a structural change with memory activity, at
the molecular level of neural structure. All such accounts, at most,
provide a hypothesis for the plasticity of brain or neural structure, or of
neural development, in its interaction with perceptual experience (of
differing environments, or of differentially deprived sensory contexts, or
of different modes of selectivity, focus and attention, in perception).
Such a physiological-structural hypothesis opens the path to research on
the effects of varying historical modes of human praxis, or of active
experience on the physiological apparatus itself, or its variability in
ontogenetic development. It does not yet provide either a model for, nor
a mechansim for *historical* changes in modes of perceptual action. That
is not because such a hypothetical neurophysiological model cannot be
conceived; rather, it is because the problem has not been posed in these
terms, and therefore, the requirement or demand for such a model is
simply lacking in current scientific research.

An historical theory of perception, on the other hand, must begin with
the genesis of *human* perceptual activity (as distinct from animal
perception) in historical praxis itself. It must begin, therefore, where
evolutionary and neurophysiological accounts of perception leave off,
accepting the terminus of the biological evolution or phylogenesis of the
perceptual apparatus as no more than the precondition for historical
perceptual development; and accepting the ontogenetic or developmental
plasticity of the neurophysiological basis of perception as only a
promissory note on some future physiological model of historical
perception.

What is this 'historical human praxis' which is proposed here as the
genesis of human perception? It is, in the first place, the fundamental
activity of producing and reproducing the conditions of species existence,
or survival. What is distinctively *human* about this activity (since *all*
species fall under this injunction of reproducing the species life) is that
human beings do this by means of the creation of artifacts. Their
production, as distinct from the foraging, scavenging or hunting activity
of other animals, proceeds by a transformation of part of the
environment into an extension of the animal organs – as, e.g. *tools* are.
But, in more generic terms, the 'tool' may be *any* artifact created for the

purpose of successful production and reproduction of the means of existence. Therefore, the use of language for communication in this enterprise makes language itself such an artifact, or 'tool'; so too is the mode of social organization, or of division of labor which is instrumental in the successful satisfaction of existence needs, or of the needs to reproduce the existence of the species. Extending the notion of 'artifact' as 'tool' still further, the acquistion of skills, in the processes of production (even at the level of foraging, scavenging or hunting, and prior to the introduction of agriculture or the domestication of animals) creates such skills as themselves 'artifacts', even where these skills do not entail the use of tools in the ordinary sense, but only the mastery of the natural organs of the body, and of perceptual skills in pattern-or-cue-recognition, for the purposes of satisfying productive or reproductive needs. The crucial character of the human artifact is that its production, its use, and the attainment of skill in these, can be transmitted, and thus preserved within a social group, and through time, from one generation to the next. The symbolic communication of such skills in the production, reproduction and use of artifacts – i.e. the teaching or transmission of such skills is the context in which *minicry or the imitation of an action becomes a characteristic human mode of activity*. It is, in effect, this ability to *represent an action* by symbolic means which generates a distinctive class of artifacts, which we may call *representations*. (2) This, then, is the second case in which perception is related to historical modes of action, in what I have called the *reflexive* sense. Such representations, then, are *reflexive* embodiments of forms of action or praxis, in the sense that they are symbolic externalizations or objectifications of such modes of action – 'reflections' of them, according to some convention, and therefore understood as images of such forms of action – or, if you like, pictures or models of them. Though I will discuss this further, later, it will be sufficient here to characterize such representations in the broadest way as the distinctive artifacts created for the purpose of *preserving* and *transmitting* skills, in the production and use of 'primary' artifacts (e.g. tools, modes of social organization, bodily skills and technical skills in the use of tools). The modes of this representation may be gestural, or oral (linguistic or musical) or visual, but obviously such that they may be communicated in one or more sense-modalities; such, in short, that they may be perceived. We may call such representations, then, 'secondary' artifacts by contrast to the primary ones (e.g. axes, clubs, needles, bowls, etc). At this stage,

the notion of 'representation' is a functional one, in the sense that *anything* which is capable of preserving and transmitting a mode of action, thus 'representing' it, serves that function. Furthermore, such representations are actual physical and perceptual embodiments of a mode of action or praxis: either in the more permanent forms of physical objects of a certain configuration, or in a certain arrangement – e.g. 'prototype' tools taken as models to be copied; visual symbols or marks engraved or painted or drawn; etc. – or the more transient forms of bodily gesture, ritual performance, utterance – e.g. in hunting rituals, dances, chants or songs. The mimetic character of such representations consists not simply in their imitation of natural objects or animals, but in their imitation and representation of modes of action or praxis. Therefore, the element of convention in the representation comes to play a large role, and any notion of 'natural resemblance' becomes too weak to encompass this range of mimesis. But such representations, as 'secondary artifacts', are not 'in the mind', as mental entities. They are the products of direct outward action, the transformations of natural materials, or the disposition or arrangement of bodily actions (e.g. in dance) or the social forms of organization of such activites as hunting, or of such relations as kinship, hierarchy, etc. They are externally embodied representations.

This excursus, broadly characterizing historical human praxis in its genesis, is meant to provide, at the same time, the context for distinctively human perception. In summary, then, what constitutes a distinctively human form of action is the creation and use of artifacts, as tools, in the production of the means of existence and in the reproduction of the species. *Primary* artifacts are those directly used in this production; *secondary* artifacts are those used in the preservation and transmission of the acquired skills or modes of action or praxis by which this production is carried out. Secondary artifacts are therefore *representations* of such modes of action, and in this sense are *mimetic,* not simply of the *objects* of an environment which are of interest or use in this production, but of these objects as they are acted upon, or of the mode of operation or action involving such objects. Canons of representation, therefore, have a large element of convention, corresponding to the change or evolution of different forms of action or *praxis,* and thus cannot be reduced to some simple notion of 'natural' semblance or resemblance. Nature, or the world becomes a world-for-us, in this process, by the mediation of such representations, (or more broadly,

such canons of representation), and thereby, in accordance with our varying modes of practice.

The purpose of this construal of human action or praxis is to set the context for the genesis of distinctively human perception, and for the changes in perceptual mode as functions of historical changes in human praxis.

4. A SCHEMA OF THE RELATIONS OF PERCEPTION TO PRAXIS

It may be useful to represent the relations of perceptual activity to other modes of *praxis* or human action, in the form of a schema. This schema places perception in a feedback loop, and proposes that it is mediated, or conditioned by the fundamental modes of praxis – production and communication – and by the instrumentalities or artifacts by means of which this *praxis* is carried out.

The fundamental division of human praxis into 'making' (production) and 'doing' (communication) follows Aristotle, who distinguished the making of things from 'human action', in the sense of relations between people. Thus, Aristotle characterized *Ethics* (from its etymological root) as that science which had to do with the social relations among people. I have broadly characterized this as 'communication' (also suggesting an etymological connotation) in the schema; though the relations *between* things to be made (*factibilia*) and actions to be done (*agibilia*) are to be understood as close. For it is in the social praxis of the production of the means of existence, and also in the conditions for the reproduction of species-life, in the sexual and social processes of generating, rearing and socializing the young, that communication itself develops in a distinctively human way. So it is the Aristotelian schema as it is profoundly transformed by Marx and Engels, in their sketch of an historical materialism, that provides the basic model here.

The means or instrumentalities of this productive and communicative praxis are the artifacts and the skills in the use of these artifacts which distinguish human praxis from animal behavior. These are, primarily, *tools* and the various modes of symbolic communication, or '*languages*', in a ramified sense. (Parenthetically, animals do provide instances of such 'praxis' as well, in primitive or proto-artifacts, and in proto-language; and therefore, I do not insist on an absolute breach between animals and humans in this regard. But incipient modes of such 'praxis' among animals are still to be sharply distinguished from the

"OFF LINE" ◄──────── PERCEPTION ◄──── REPRESENTATIONS◄──┐
⎡ Contemplation ⎤   (images, models)
⎢   Aesthetic or ⎥   ⎡ Language ⎤
⎢ "Disinterested" ⎥   ⎢ Pictures ⎥
⎣   Perception   ⎦   ⎢ Dance   ⎥
                                                                ⎣ Poetry ⎦

                                (Doing)   Means of
                             COMMUNICATION ──────────►   Communication

  └──► HUMAN ACTION   (skills) ──► ARTIFACTS
     ►   or PRAXIS

                             PRODUCTION ──────────► Means of
                              (Making)   Production

                                                       TOOLS ◄──

"OFF LINE" ◄──────── PERCEPTION ◄──── REPRESENTATIONS
⎡ Planning   ⎤   (Prototypes, Modes
⎢ Rehearsing ⎥   of Reproduction)
⎣ Practising ⎦

dominant mode of this activity among human beings, and therefore, from the cultural transmission and cultural evolution, which the use of tools and of language makes possible. I take the artifacts (tools and languages) to be objectifications of human needs and intentions; i.e. as *already* invested with cognitive and affective content. The tool is *understood*, both in its use, and in its production, in an instrumental fashion, as something to be made for and used for a certain end. What is characteristic of the production and reproduction of tools, then, is the conscious teleology of that productive praxis; just as the *use* of a tool, obviously, already embodies this intentionality as well. More radically, I would claim that it is in the acquisition of these skills that intentionality or conscious teleology makes its *first* appearance in the world. The counterargument, of the anti-teleologist, might be the claim that such skills or modes of goal-oriented behavior are simply conditioned responses, or habituation, shaped by (natural) schedules of reinforcement; and that the attribution of both 'consciousness' and 'teleology' is simply an unnecessary multiplication of entities, loading a specifically adaptive mode of behavior with too much cognitive baggage. My argument is that it is precisely the evolution of representation, or of symbolic embodiments or objectifications of modes of action or praxis,

in an objective artifact, that provides the very genesis of such cognitive consciousness and of such teleology. It is in the use of such representations that a characteristic mode of praxis is preserved, and comes to be transmitted; and in this lies the germ of cultural evolution, substituting the artifact, and the adaptive changes in the modes of social-historical praxis for the alternative and slower *genetic* means of preservation and transmission which function at the biological level. (Thus, as I say in another paper, the *artifact* is to cultural evolution what the *gene* is to biological evolution.)

The objectification of human intention is embodied both in the tools used in production, in the skills acquired and adapted to this use, and in the forms of symbolic communication which develop in language, in art, in dance and poetry, in their origins. Now, it is my argument that our perceptual activity is an activity mediated or conditioned by these very forms themselves. Insofar as our seeing, hearing, etc. are themselves modes of action, the dominant forms of representation are the filters of the purely biological perrceptual mechanisms; and more than this, actually transform the function (and speculatively, also the structure) of these mechanisms. Thus, in the schema, the very use of tools for a certain purpose is what determines how such tools will be seen, handled, etc. The usual view is that this 'understanding' is super-added, as 'interpretation', upon a purely physiological or psychological base (of sensory input; or of 'pure' or 'pre-interpretive' perception). The 'seeing as' arguments refine this view, taking the context as always determining or influencing the way in which something is perceived, so that, at the limit, all 'seeing' is 'seeing as'. But even this notion – e.g. in the discussions of the theory-ladenness of observation – have a residual notion of the *given*, as that which *then* becomes laden with theoretical or other background-knowledge context. What is wrong with this is similar to what is wrong with the notion of 'state of nature' in classical political theory. It is (as Rousseau already recognized) a philosophical fiction, invented for theoretical purposes, and the mistake is to assume that it has existential import as well – i.e. that a 'state of nature', or an 'innocent eye' exists somewhere at the *basis* of what then gets added to it. But if the very origins of the perceptual system are seen in the interactive adaptation of organism and needs to environment, there is no 'beginning' at which such perceptual neutrality, or a perceptual 'state of nature' exists. 'Nature' is always the arena of activity, and for perceiving organisms, what is 'there' or 'given' is always a product of their activity and that

cross-section of the world which this activity encounters and transforms into an environment. In short, 'environment' is itself not a neutral term, but is what is functionally adapted to, and changed by an organism, or by a population of organisms, (some of which then come to constitute parts of the environment for others). One may paraphrase Marx, in this ecological context, by speaking of an "ensemble of natural relations".

In the human case, the transformation is by means of artifacts. But the very production and reproduction of artifacts presents a made world as the symbol of what there is, and as a representation of the modes of praxis themselves. The tilled field, or the domesticated animal is no less an artifact in this sense than is the spear or bow or pot. Moreover, the very environment itself, as a space of action, is invested with the characteristics of an artifact. Nature becomes transformed, not only in the direct practical way of becoming cultivated, or shaped into objects of use, in the embodied artifacts we call tools, or in the 'instruments' of existence such as clothing, houses, etc.; it becomes transformed as an object or arena of action, so that the forest or the river itself becomes an 'artifact' in this ramified sense. It is a source of food, or of danger; it has a direction or is mapped into regions; it is endowed with familiar properties, or anthropomorphized in a *representation* of its uses, and of the modes of praxis appropriate to such uses. To take an extreme example, perhaps: the hunter, hearing a crack of a branch, or seeing a sudden flight of birds, transforms that very sound and sight into an artifact – an instrument – of the hunt itself. But what the cracking branch is heard *as* is already an index of a social mode of praxis – of hunting, in this case – and therefore, insofar as the hunt has a specific historical form or function, nature itself has become historicized and socialized, and has come to be a *representation* of a certain mode of praxis or human action.

All the more so, if this rather extreme version of perception by means of an 'artifact' of the hunt is translated into the more acceptable modes of representation: the actual pictures of animals, and of hunting; the actual mimetic rehearsal of the hunt in ritual dances; the actual telling of a story in which actions are portrayed in a certain sequence, outcomes are described and the lesson or moral drawn.

That we come to see by way of our picturing I have argued in an earlier paper. 'Pictures, Representation and the Understanding' where it was shown that our canons of representation, the very styles and conventions of our picturing teach us to see differently. The artist, in effect,

reeducates us perceptually, when he presents us with a possible world, different from that which is culturally dominant at a given time. The plasticity of our forms of perception is evidenced, for me, by the fact that, as styles or canons of representation change, historically, the world has seen changes as well. Possible worlds become actual, differentially. It is also a fact that we can live, perceptually, in alternative worlds, adopting different 'sets' or different canons of representation for different contexts.

The schema proposes that the forms of representation – (either in the symbolic function of tools and skills of action as themselves models of the modes of action involved in their production, reproduction and use; or in the use of representational language, pictures, mimetic performances) – become the very conditions of our perceptual understanding. The feedback loop concerns the way in which a particular mode of perception, mediated by such representational forms, is itself involved in our productive and communicative praxis. What we have learned to see something as, becomes in turn, the guide to our outward practical activity. Just as, at another level, the perceptual constancies (e.g. for shape, size, distance) are crucial in orienting the motion and activity of animals in their life space, so too in the ramified cultural life-space of human praxis, the adaptations and changes of perceptual mode become crucial in orienting the forms of human action.

This feedback loop of changing praxis and changing perceptual modes has a relatively autonomous outer loop as well, which I have characterized in the schema as 'off-line'. But it is all-important. Here, the forms of representation themselves come to constitute a 'world' (or 'worlds') of imaginative praxis. The mimetic reenactment of the hunt is not a hunt: no animal is killed, and no hunter is endangered. In this sense, the rehearsal for the 'real thing' is 'off-line'; and indeed, as such an imaginative reenactment, it can come to achieve a greater or lesser distance from the performance itself, in that it allows for conventional elaboration, and for the investment of values and needs of a related sort – e.g. for communal participation, for purposes of courtship, or of kinship display, for the expression of joy, determination, aggression – all of which may be indirectly related to the hunt. Still, this relative detachment from the actual direct praxis of the hunt, and the symbolic plasticity of the ritual form itself comes to make the 'off-line' representation a receptacle for the expression and reenactment of a wider range of cognitive and affective needs, intentions and values than if it

were merely or only a hunting-rehearsal. On this reconstruction, we may
speak of a class of artifacts which can come to constitute a relatively
autonomous 'world', in which the rules, conventions and outcomes no
longer appear directly practical, or which, indeed, seem to constitute an
arena of non-practical, or 'free' play or game activity. This is particularly
true when the conventions of representation – e.g. in art, or in
language – become transparent, i.e. when the relation to direct
productive or communicative praxis is so weakened, that the *formal*
structures of the representation are taken in their own right as primary,
and are abstracted from their use in productive praxis. So called
'disinterested' perception, or aesthetic perception, or sheer contemplation
then becomes a possibility; but not in the sense that it has *no* use. Rather,
in the sense that the original role of the representation has been, so to
speak, suspended or bracketed.

   An alternative theory for the source of this purely formal or
disinterested perceptual activity is that, apart from the sheer utilities of
productive praxis, the organism (higher animals and humans) has a need
for the spontaneous play activity, in which its faculties are exercised
beyond the limits of present need. In Groos' theory, in *The Play of
Animals*, he proposes this as 'premonitory' activity, in which the young
animal does not *imitate* adult behavior, but rather practises, in an
instinctive way, those activities – (e.g. fighting, hunting, sex-play,
etc.) – for which the need lies only in the future. This is presumably an
adaptively selected 'play-instinct' already conditioned by the species-
purposes which it serves. There may, in fact, be, in humans as well, a
need, rooted both in biological and in socially evolved contexts, for such
'free' activity; and its relation to the directed forms of 'necessary' praxis
is surely complex. But I will do no more than suggest here that this in no
way affects the general thesis that such disinterested or 'off-line' activity
depends in its *formal* structures on the practical rules, rituals and modes
of praxis which are represented in the 'on-line' models of this activity.
Which is initiatory is an open question; my own view, at present, is that it
is the direct forms of necessary productive praxis that generate the
representational forms themselves; and that only by this means is the
perceptual activity mediated and does it become historical. The artifacts
of the imaginative construction of 'off-line' worlds I take to be
derivative, and abstractive. But there may well be a structural component
in all this which derives from other (though no less social) needs which
transcend the more immediate necessities of productive praxis.

I would characterize such artifacts, abstracted from their direct representational function, as 'tertiary' artifacts, and suggest that they constitute a domain in which there is a free construction in the imagination of rules and operations different from those adopted for ordinary 'this-worldly' praxis. Such possible worlds may indeed reflect the limits of the perceptual praxis in a given 'actual' world – i.e. a world in which direct outward and necessary productive praxis takes place, in accordance with rules, and ontologies evolved through this praxis. That is to say, just as in dreams our imagery is derived from our ordinary perception, but transcends or violates the usual constraints, so too in imaginative praxis, the perceptual modes are dervied from and related to a given historical mode of perception, but are no longer bound to it.

Yet, the feedback here is important. If, as I claim, an 'actual' world is a historically selected or achieved one, corresponding to a given level of social-historical praxis, (technology, social organization, etc.), then the 'possible' worlds provide candidates for conceivable change in this mode of praxis itself. The perceptual alternative provides the possibility of a practical alternative, as, so to speak, a perceptual hypothesis. Such imaginary worlds I do not take as 'dreams' or 'in the head', but as embodied representations, or better, embodied alternative *canons* of representation: embodied *in* actual artifacts, which express or picture this alternative perceptual mode. Once the visual picture can be 'lived in', perceptually, it can also come to color and change our perception of the 'actual' world, as envisioning possibilities in it not presently recognized.

The activity of the imagination is therefore a mode of alternative perceptual praxis, and is 'off-line' *only* relative to a historically actual or dominant present mode of perceptual praxis. What the imagination is, as 'internal representation', i.e. as a picturing 'in the mind' of such alternatives, I take to be derivative from the actual making of imaginative artifacts. That is to say, in its genesis I take imaginative praxis to be praxis in the actual world, or the actual production of representations; the interiorization of these representations, as 'mental' artifacts, I take to be a derivative process. In this sketch, I cannot say more than this, programmatically, and it remains to be elaborated in a subsequent paper.

The upshot, however, is that the constructions of alternative imaginative perceptual modes, freed from the direct representation of ongoing forms of action, and relatively autonomous in this sense, feeds back into actual praxis, as a representation of possibilities which go beyond present actualities.

That this *is* a mode of perceptual activity, and not simply some abstracted mental imaging should be clear from the very nature of imagery itself, insofar as it is both derived from, and in turn helps to shape and inform ordinary perceiving. The operations of anticipation, familiarity, resemblance, of socalled *'Einstellung'* or 'set' in perception; the selectivity and focus of perception, its involvement with needs, intentions, and feelings, with cognitive and theoretical frameworks, all speak to the inseparability of perception from the whole ensemble of social and individual relations in which it functions, and of which it is an expression. Perceiving is therefore not an *incipient* form of human action; it *is* human action in one of its modes, complexly and subtly involved in all the other modes of more direct productive praxis, or in the motor-activity by which human beings act in the world, and sustain their existence. My argument has been that, because of this thorough integration of perception with praxis, its forms change historically as that praxis changes historically; and that it is both determined by and helps to determine these very changes themselves. An historical epistemology therefore undertakes the task, (sketched here only programmatically) of investigating both the mechanisms of change of perceptual modes, and the history of these changes. Such a history of human perception cannot be an abstract philosophical enterprise, though it relies on the analytic and scientific investigation of the logic and process of perceptual activity, the physiology of human perception, and the characteristic experimental study of perceptual phenomena. In addition to this, however, a history of perception needs to investigate the historical changes in social praxis, and in the 'world' of artifacts and modes of representation which impinge on perceptual change. In this sense, the history of technology, of science and of art become relevant contexts; and so too does that social and cognitive history of human praxis which involves the contexts of belief and ideology, of world views and value-systems – in short, of that social human world in which perception has its genesis, and in which it functions.

RULES AND REPRESENTATION: THE VIRTUES OF CONSTANCY AND FIDELITY PUT IN PERSPECTIVE

[1978]

In this paper I will argue that the widely accepted theory of perceptual constancy, and the equally widely held account of fidelity in representation rest on the same mistake. My argument (which derives from and extends Nelson Goodman's, in *Languages of Art*) is that this theory of perceptual constancy is based on a theory of vision which is false, namely, the standard theory which interprets vision on the model of Euclidean geometrical optics. Further, I will argue that the view which takes the rules of linear perspective to be the norm for fidelity in pictorial representation is based on the same false theory (or on its mirror image, that is, on the theory that pictorial representation's norm for fidelity is the mirror image).

My argument, in brief, is that the theory of perceptual constancy is required to explain perceptual invariance of such features as size, shape, etc., *only* because it construes our perception of objects and scenes in the three-dimensional world as if it were perception of a two-dimensional image of the world. I hope to show that the theory of constancy is redundant: that it is required only in order to correct for putative mistakes which the theory itself mistakenly postulates as features of our perception. The theory of perceptual constancy thus invents the vice with respect to which it is the complementary virtue. In effect, I hope to show that the theory cancels itself out, and is therefore redundant, once rightly viewed. Moreover, since the account of representational fidelity as defined by the rules of linear perspective is interdependent with constancy theory, this notion of fidelity also stands or falls with the theory. To give way to metaphor for the moment, constancy and fidelity are parallel virtues. Put in perspective, they will be seen to converge. Let me begin to put matters in perspective then, first, by considering the standard view, and then, by presenting my argument against it.

The theory of perceptual constancy tells us that our perception preserves such features as size and shape invariant through changing appearances. The rules of linear perspective, in drawing, tell us how to render

these changing appearances with fidelity so that constancy can do its work on two-dimensional pictures just in the same way that it does on three-dimensional objects and scenes. Constancy and fidelity are held to be virtues in that they give us the rules for correct perceptual interpretation and correct representation, and thus permit us to see things as they really are, and to represent them as they really appear, without error. Constancy preserves reality, while fidelity preserves appearances. They are, in effect, inverse virtues. What constancy corrects, according to rule, is just what fidelity presents for such correction.

The virtue of constancy, in perception, is that it corrects or regulates putative perceptual mistakes. Or so it is alleged. The theory of constancy is thus a normative theory, in the face of what, without it, would result in perceptual error. Thus, shape or size constancy is the perceptual system's way of setting right what is presented in a misleading way. Tilted circles 'appear' or are presented to our visual apparatus as ellipses. Constancy stretches them back into shape, so that we recognize the presented elliptical shapes as tilted circles, as fully round as circles ought to be. Things in the distance 'appear' smaller, or project a smaller retinal image than do the same things closer up. Constancy preserves size through such phenomenal or physiological variations, and we are said to 'infer' variations in distance, as the correct reading of apparent size variation.

Fidelity is the virtue of constancy transferred from three-dimensional vision to two-dimensional pictorial representation. Fidelity, thus, is a virtue in representation which preserves in the appearances what constancy preserves in ordinary perception. If perceptual constancy is the perceptual system's rule for right readings, then fidelity is assured by the rule of representation which delivers up the text for the right reading according to the perceptual rule. Or so it seems.

This abbreviated version of the traditional view of perceptual constancy and representational fidelity raises some conceptual questions, which it may be useful to discuss at the outset. First, a number of brief clarifications and qualifications:

(a) The whole account of perceptual constancy presupposes a distinction between what is *received* by the visual apparatus, and what is *perceived*. The classical distinction between sensation and perception, between the 'givens' in perception and our perceptual 'judgments,' is crucial to the notion of constancy. What is *received* is ostensibly the

bundle of light rays reflected from surfaces. The external medium (air) is presumed to have no effect on the geometry of the projection, so that in linear perspective at least, the conditions of Euclidean space are preserved. (In aerial perspective, where color and dark-light gradients are affected by distance, matters are of course different.) In the projection of the bundle of light rays, variations in size and shape of projected figures follow the transformations of the Euclidean geometry i.e. of classical geometrical optics. Whatever refraction there is, is similarly described in terms of classical dioptrics, so that lenses, for example (or more generally, media with differential refractive indices) bend the received light rays. And in the case of the crystalline biconvex lens of the eye, the flux of light rays is converged, and at the retinal distance from the lens, forms an inverted image, or point-for-point projection on the retinal surface. This image varies in size, inversely with the distance between eye and object; and varies in shape with variation in the angle of incidence of the light rays with the plane upon which the image is projected. [1]

The distinction therefore is one between the optical image on the retina and the object or scene of which it is the projected image. More precisely, the distinction is one between certain features or properties of the image and of the perceived object, e.g. size and shape. The presumption that a corrective operation is required in visual perception is based on this difference: for the 'given', or the stimulus-information is taken to be the retinal image itself. Once this premise is accepted, then of course the preservation of the shape or size of perceived objects through variations in the retinal image requires some mental or psychological operation – what Helmholtz called '*unbewusster Schluss*,' an unconscious inference. The fundamental assumption of the constancy-theory is therefore that our access to the three-dimensional visual world is mediated by a two-dimensional image, which is a projective transformation of the bundle of light rays, by convergence, upon a two-dimensional surface. It is easy to understand why this model lends itself easily to the interpretation of perception as a mode of reflection upon an internal 'picture' of the world; and therefore, that the perception of the world is like picture-perception, in some sense. [2]

But what I take here to be the error of constancy-theory – (the interpretation of the binocular three-dimensional vision of a moving and acting subject *as if* it were the monocular two-dimensional vision of a

fixed subject) – is an instructive error. For constancy theory, though it is a false theory of vision, turns out on a different interpretation, to suggest an alternative theory, like my own. Namely, if the constancy transformation is required, by the theory, to preserve invariances through 'phenomenal' or physiological (retinal) variation in what is 'given', or presented to the eye; and if what is given is a two-dimensional mapping or 'picture' of the visual world, then the implicit claim is that our seeing is mediated by such a 'picture.' My own argument is just that such a 'picture' does mediate our vision, but that it is not the 'picture' *given* in sensation, or by the retina. Rather, it is the actual mode of picturing which we engage in, in *making* pictorial representations, which performs this mediating function, and to which (different) constancy theories are therefore relevant (depending on different rules of transformation in representation.) Construed in this way, constancy theory is not a theory of visual perception *per se*, but really (and unwittingly) a theory of *picture*-perception! It follows, therefore, that *a given* theory of constancy *defines* a criterion of fidelity, as its concomitant, namely just that rule of perspective representation which presents the group of transformations characterized by the invariance which the (particular) constancy theory proposes.

The group of transformations which *defines* linear perspective as the norm of fidelity in representation is just that of Euclidean geometry (as a *formal* mathematical system) and its interpretation for light rays, reflection, and refraction, in Euclidean geometrical optics. My argument is therefore that it is the theory of vision which interprets the visual system in terms of Euclidean geometrical optics which serves as the *theoretical* warrant for the fidelity of perspective representation. This theory therefore not only organizes the empirical evidence in a given way, but also sets the framework for what will be *taken* as evidence, and what will count as experimental data. It therefore organizes inquiry on the model of the theory. For example, the theory *defines* light-rays as straight lines, and defines image formation through projective transformations in terms of Euclidean point-for-point mappings of the intersections of such projected rays with plane surfaces. If the experimental evidence shows that such a mapping does *not* take place in the physiological projection on the retina, for example, then the theory is false. But in fact, the retina is not an undifferentiated plane surface. It is a differentiated curved surface (differentiated by the distribution of various arrays of receptor-neurons,

e.g. for slant, edge, color, etc.).[3] The theory is therefore false on these grounds alone. But it is also phenomenologically false, if it is taken not simply as a theory of the physical (or physiological) geometry of vision, but as an explanation of what we see. For pictorial representation in accordance with strict linear perspective *looks* wrong.[4] Now this is a tricky point. For if *I* claim that a picture 'looks wrong' when it is made in accordance with an unmodified rule of Euclidean projection, then it would seem I am appealing to some pre-pictorial criterion of visual fidelity – e.g. perception *per se* – and this would be in contradiction to my hypothesis that our vision is mediated by representation. Here, instead, it would seem that unmediated vision becomes the test of fidelity. The answer to this objection lies, I think, in the fact that our vision is never simply the product of a *given* norm of representation, but is a complex process mediated by a group of norms, some deriving more directly from our biological and practical activity (e.g. from the physiological basis of our visual system, or from the forms or structures of our motor-activity, or our non-pictorial *praxis*) and some from the *different* norms of representation which have developed historically, and which form, so to speak, our visual heritage, or parentage. The fact that painters have, from the start, modified the rules of strictly geometrical linear perspective bespeaks a certain autonomy in the choice of norms, and in the modification and elaboration of norms in our representational practice itself. But this is the subject for a different inquiry.

My argument against the standard view is presented in the context of Nelson Goodman's discussion of perspective in *Languages of Art*. Goodman has presented the most striking and challenging argument against the standard view.

In the brief section on perspective in *Languages of Art* (pp. 10–19), Nelson Goodman argues that "Pictures in perspective, like any others, have to be read; and the ability to read has to be acquired." (p. 14) Against the standard argument that perspective representation duplicates a bundle of light rays reflected from the object itself, under specified conditions, Goodman argues that the "specified conditions" (e.g. fixated monocular vision through a peephole, frontal-parallel presentation, at a certain distance, etc.) are so "grossly abnormal" that "to measure fidelity in terms of rays directed at a closed eye would be no more absurd." (p. 13) In short, Goodman's argument is that the claim that two-dimensional

perspective representation delivers the same visual information as that of the three-dimensional objects represented, is based on such a contrived mode of presentation, and such an impoverished and abstracted mode of vision, that the 'likeness' or 'fidelity' of image to object is itself a construct no less subject to rules of representation than any other *non*-perspectival representation. It is just that the rules are different.

Now the reason such an argument as Goodman offers seems perverse is that our own familiar canon of fidelity in representation is that of linear perspective. I have argued, (in 'Pictures, Representation and the Understanding,'[5]) that the adoption of this canon is an historical act, which involves the adoption and interpretation of Euclidean geometrical optics as a theory of vision; and that this theory is false. Further, that our seeing is itself not simply a physiological, but a social and cultural activity, and that our adopted modes of representation guide our seeing itself. "We see by way of our picturing," I said there; and meant by it that we come to 'see' tilted circles as ellipses only because we have come to represent them as ellipses in our adoption of perspective as a canon of representation. In fact, we continue to see tilted circles as tilted circles *except* when we are asked to represent them; and then we represent them as ellipses in order to make them look like tilted circles suitably drawn in accordance with the rules of perspective (or the theory of geometrical optics as a theory of projection of reflected light rays upon the retina).

Here, I should like to present an even sharper and more perverse sounding argument, which follows, I believe, from Goodman's account; but which offers a radical reinterpretation of the phenomena of visual constancy – in particular, shape and size constancy.* It is this: that the very discovery or description of the phenomena of constancy cannot make sense except as an interpretation according to the rules of perspective (or the theory of vision based on geometrical optics). And that, in fact, the very allegation that we 'preseve constancy' through projective transformation is not an account of the way the eye or our visual system works, but is rather dependent on the adoption of a false theory of vision,

---

* I deal in this paper only with shape and size constancy. Whether my argument can be interpreted for object or color constancies, or for others, I do not know as yet. But it would be a different argument, since what is at issue here is the perspective transformation of size and shape, and as yet, nothing more.

linked to a convention of representation (which is itself neither true nor false).

The standard view, for example, holds that parallel lines 'visually' converge, when presented in anything but the frontal-parallel plane (and even there, as their distance above or below the line of sight increases). Thus, railroad tracks appear to converge in the distance. By the adjustment we make (e.g. by 'unconscious inference,' in Helmholtz's classical phrase), we 'correct' for the *apparent* visual convergence, and interpret the visual (i.e. retinal) image of converging lines as parallel, but as going off into the distance.

My argument is that we do not make any 'correction' or 'inference' at all; but that parallel lines going off into the distance appear, in normal binocular vision, to be just what they are – parallel lines going off into the distance, *without* convergence. Constancy does not need to be preserved. It is given. Yet, all of us *can* see parallel lines as converging in the distance. That is to say, we can willfully violate the given constancy. And I claim that we can do so only because we have adopted a perspective mode of representing parallel lines as converging, in our pictorial, or two-dimensional representation of such lines going off into the distance. But why should we adopt such a mode of representation? The standard argument says that this reproduces or matches the visual information delivered to the eye, i.e. it reproduces the retinal projection of the reflected bundle of light rays from the objects themselves; and that this in fact can be measured. Goodman's argument, I take it, counts against that, and I will assume it here, for the moment, in order to go further. But I will return to it later, in the light of recent psychological studies of picture perception, and in particular J. J. Gibson's alternative account. Why do we, then, come to adopt the perspective representation of converging lines as canonical? The answer requires that we play some tricks with mirrors.

Mirrors are strange, because deceptive. If we strip ourselves of the historical familiarity with mirror images for the moment – i.e. 'bracket' our easy accommodation of the mirror image – then we may recapture the magic of the mirror: it delivers a three-dimensional image from a two-dimensional surface. We discover the mistake of identifying a mirror image with the object imaged when we bump into the mirror. Kept from bumping, and from surround-cues (frame, blemishes or distortions or

glossiness or glaze-effects which reveal the mirror surface, etc.), we may be fully deceived. In fact, on Goodman's analysis of metaphor, by some small stretching, mirror images may be seen as metaphors, in that any image taken for the object imaged is a label, i.e. a non-verbal visual metaphor, taken for something it is not. The mirror image is a peculiar 'metaphor,' however, in that it is hardly a "calculated category mistake" as Goodman says of verbal metaphors (p. 73).

In any case, it is not true that mirrors do not lie. They assuredly do, when we do not recognize the images in them (or 'on' them) as mirror images; that is, when we mistake what appears on (or from) a two-dimensional surface as a three-dimensional object. Recognizing the mirror as a mirror and the image as an image requires some small sophistication.

Now add to this small sophistication the rather grand move of interpreting the eye as a mirror of objects, and of doing so not only naively, but theoretically. That is, consider that the mathematical theory of Euclidean projective geometry, suitably interpreted as the classical theory of optics (with rectilinear propagation of rays, reflection and refraction of light rays from plane or curved surfaces, and convergence of such rays through pinholes or lenses) is taken as the theoretical account of image-formation. The result is that the two-dimensional mirror image will be taken to be the same sort of image as that projected on the 'mirror' of the eye. And if the mirror image is itself to be represented, that representation will be formed by duplicating the surface image on the mirror. Literally, the mirror image can be transferred to a pictorial representation either by tracing it directly (e.g. with a crayon on the mirror's surface) or by projecting such an image through a lens or a pinhole onto a plane surface, and tracing it there; or, as Leonardo Da Vinci proposed, tracing it on a transparent pane of glass held before the eye. This is, in effect, what the *camera obscura* provided for the Renaissance. Thus, Brunelleschi's injunction to "draw it the way it looks in the mirror" is the craftsman's practical version of incorporating the laws of geometrical optics into the rules of pictorial representation.

Plainly, when the parallel lines going off into the distance are mirrored or projected on a plane surface, they *do* converge; and so they are represented, in perspective drawing. But my argument is that they do not appear to converge in three-dimensional binocular vision, *except* when

we have learned to transfer the convergence adopted in perspective representation to our actual *abstractive* vision of the three-dimensional world. Parallel lines appear to converge only after we have learned to see by way of our picturing, and do so only when we are able to translate our vision into the language of our representation. That is to say, the historical adoption of a rule of representation affects our perception itself, to the degree that we can now 'see' the convergence when we put ourselves into the framework of the new rule. The broader argument, which I develop elsewhere, is that our modes of perception change historically, in accordance with changes in the modes of our social or cultural practice – in this case, the practice of pictorial representation. This is an aspect of a general view which I call historical epistemology.[6]

What does this do to the usual account of perceptual constancy in this case? I would argue that perceptual constancy can be introduced as a concept, and indeed, becomes an actual perceptual phenomenon, only when what has to be 'corrected' for is our pictorial mode of perspective representation formed in accordance with a theory of vision based on Euclidean geometrical optics.

Now this is a fairly strong claim. In its strongest version, it alleges that there is no visual processing of variable imputs which preserves constancy, and that, in our normal vision, none is needed. This would be tantamount to saying that naive vision is correct, and thus needs no correction by a constancy transformation. But of course, there is no such thing as naive vision. The very evolution of the eye as an adapted visual mechanism embodies those processes and functions which vision has to serve. Since, naively speaking, we want to account for the visual ability to recognize the *same* object in variable presentation, (i.e. through variations in angle of sight, distance, light conditions, surround, etc.), we have to build into our theory of vision an account of how such identification is possible. But the demand can be made only if this identity is put in question by variation. And to the 'naive' (i.e. highly complex, adapted, evolved) eye, it is *not* put in question. For the 'naive' eye is evolved precisely to preserve this identity. At most, it would seem, object and shape constancy, if they are not innate, are acquired very early, among human infants,[7] and seem to be innate for criterial and functional objects and shapes among animals.[8] What puts constancy in question, then, is our own sophistication, and our science. We are able to abstract from our

visual activity; but only because we have acquired an ability to represent three-dimensional objects and scenes visually on a two-dimensional surface; and because we have learned to construct theories of vision based on this activity. It is the abstractive *practice* of representation which introduces the possibility of visual abstraction in ordinary perception of objects. We can 'see' the variations in shape and size which objects present to us only because 'shape' and 'size' have become separable, i.e. abstractible visual concepts, by means of our practice of representation. Apparent convergence of parallel lines in the distance, diminution of size with distance, and other phenomenal variations which require the mechanism of constancy to explain the veridicality of our vision, are all phenomena we would not experience, were it not for our ability to represent things 'the way they look in the mirror.' The very notion of 'phenomenon' or of 'the appearance of things,' by contrast to the way they 'really are,' is a cognitive and perceptual act of abstraction, *not* built into the perceptual apparatus, but achieved by reflection and inquiry.

Now this is *not* to argue that we should not represent things 'the way they look in the mirror.' It is rather to acknowledge that we do, but that we do not always, and that our choice to represent them in this way is an historical choice, conditioned by particular norms adopted for particular purposes. This does not yet settle the question as to whether this choice – i.e. the choice of the rules of perspective representation – is a choice of the 'correct' rendering of our visual world. It does not settle the question because my argument puts in doubt the very way in thich the question is raised.

The very notion of 'correct rendering' presupposes not simply that there is a norm, but that among alternative norms, one ought to be adopted because it yields the 'correct rendering'. But this further presupposes that the choice of one among alternative norms of representation – e.g. of linear perspective over its alternatives – itself is determined by some norm – a metanorm, or norm of norms. As in the question of choice among alternative theories in science, this question also asks for a criterion of choice (or of acceptance or rejection). To say, as I do, that the choice is 'historical' leaves it apparently norm-less, i.e. seems to concede to an historical relativism with respect to 'correct rendering' or 'fidelity'; to a conventionalism, with respect to the nature of norms; and, at best, to a pragmatic criterion with respect to which norms suit our purposes. The

alternative would seem to be the adoption of an essentialist view of the 'norm of norms,' namely, one which takes *truth* as such a 'norm of norms' and therefore defines fidelity in representation in terms of some notion of 'the way things *really* look.' Between the Scylla of historical relativism and mere descriptivism (in which each norm is its own warrant of 'correctness' – a variant on Pirandello's "Right you are if you think you are") and the Charybdis of essentialism, there is a narrow strait which I hope to navigate. I confess that the difficulty is as great here, in the question of choice among norms of representation, as it is in the context of norms for choice among scientific theories, though it is a different question.

I will return to this question of the criteria for fidelity in representation shortly. But first, let me briefly summarize my argument thus far and raise some questions against it, in order to go further: The visual system is evolved to perceive constancies of shape and size, i.e. to see objects and scenes from different viewpoints and at different distances *without* variations in size and shape. The ability to *attribute* variation, as a result of changes in viewpoint or viewing distance, is an achieved ability, which results from the cultural and historical adoption of a particular mode of representation, e.g. that of linear perspective. The theory of perceptual constancy alleges that this ability is part of the physiology and the psychology of 'normal' human and animal vision. I am arguing by contrast, that this ability is (historically) learned, *by means of* an achieved cultural practice of representation.

In effect, then, what I am arguing for is the *inverse* of the traditional view. The traditional view – the theory of perceptual constancy – alleges that the visual system *receives* variables and *perceives* constants; that is, it constructs (by 'unconscious inference,' or by some mental processing) a veridical picture or map of the external world, which is then imposed on the variations in the information which the flux of reflected light presents. I argue, by contrast, that the visual system is already structured to perceive constants, and that the *additional* ability to perceive variations is an *achieved* one; that is, that we learn to make inferences *to* the variations in shape and size, and not *from* them; and that this ability derives from the theoretical analyses of vision, which are embodied in our canons of representation. It is therefore *because* we make pictures according to the rules of perspective, that we learn to 'see' the size and shape variations of

objects in the visual field. To put this another way: The visual field itself –
the space of our visual activity and of the human practice which involves
vision – is a construct which is ordered by our practice, in particular, by
our practice of making pictorial representations of the visual world.

Therefore, I argue, the theory of perceptual constancy is false about
vision *per se*. I should qualify this, now, and claim that the theory of
perceptual constancy holds *only for* that particular mode of visual activity
which is derived from, and dependent upon picture-perception, i.e. that
mode which already interprets the visual world as a picture, or sees the
three-dimensional visual world through the 'lenses' of a two-dimensional
pictorial representation; in the case at issue, specifically as a two-
dimensional pictorial representation made according to the rules of linear
perspective.

Now one may raise against this view at least two serious objections:
first, how could one *tell* that so-called 'naive' vision picks up constancies
without a transformation of variable inputs? What empirical basis could
one adduce for the claim that the ability to notice variation, and to
'correct' for it, is learned? *Second*, what could it mean to claim, as I do,
that the acquired constancy transformation is the result of the adoption of
a rule of representation which then guides our perception, i.e. of a norm,
and not simply the (unconscious) operation of the visual system according
to a biologically or neurophysiologically based law, or on the basis of the
physical theory of optics?

(1) To the first question, there is a systematic answer, and a developed
countertheory in the work of the psychologist, J. J. Gibson.[9] Briefly,
Gibson argues that the organism is evolved to pick up (visual) invariances
from what he calls the 'ambient light,' and that these higher-order
invariances are given, so to speak, in the stimulus-information itself.
Thus, it is not *variation* which is given, and then transformed, but rather
invariance. In Gibson's view, the visual system is to be understood
ecologically, that is to say, as the system developed for an organism which
is active, moves about in the world, and therefore sees the world in terms
of those invariances which are required for its activity. Hence, he argues,
the appropriate optics for such an ecological account of vision is not
physical optics, but what he calls ecological optics.

Gibson's theory is not in itself the empirical evidence for the view that
the ability to notice shape and size variation with changing view-points

and viewing distances is an achieved ability. But it does offer an alternative theoretical account which permits us to interpret the evidence in a different way. That is, it breaks us loose from the traditional interpretation of the evidence which constancy theory offers, concerning the 'givenness' of size and shape variation in the stimulus-information, or in the stimulus image. Methodologically, I want to claim therefore that it is not fundamentally a question of what the empirical evidence is for one view or the other, but rather a question of what will be taken to *count* as empirical evidence, and how this evidence will be interpreted. To put this matter briefly, without going into the experimental data itself: there is no reason to think that the visual stimulus, or 'image' is *originally* variable, *except* on the basis of a theory of visual image formation, or on the basis of a *practice* of representing and noting this variation. Variability in the appearances is something *we* introduce into the visual world, *upon reflection*. (My extended argument on this, given elsewhere, is that relection arises with, and is concomitant with representation.)[10] More generally, my argument is that the distinction between 'appearance' and 'reality' is a distinction which requires theory, and that without theory, the visual world offers no difference between appearance and reality: what appears is just what it appears as. Or to put it differently and pretheoretically: what is, appears as what it is. That is not to say that we are pre-theoretically limited only to phenomena, or appearances, and get 'behind them' by means of theory; but rather, that the very distinction makes no sense, and does not exist, pre-theoretically. (Naive realism, for example, is not a mistake. It is simply naive, i.e. pre-theoretical and unreflective.)

   Here, the experimental evidence is fairly clear, in fact. The more 'naive' the subject, the less influential is the effect of the norms of representation on what the subject reports as what he/she 'sees.' Now this may seem to be a vacuous claim, or a circular one, if indeed *naïveté* is defined with reference to knowledge or ignorance of the rules of perspective representation. But I am claiming no more than this. If, in fact, experimental subjects who have little or no familiarity with, or practice in the canons of perspective representation, report or exhibit that the shapes they are presented with are (more or less) invariant in differing presentations (and more or less so in proportion to their *naïveté*), then it would seem to me to follow that it is the educated vision

(e.g. of teachers of perspective drawing, in some of the experiments) which is prone to note the variation for which constancy is the 'correction.'[11]

Gibson's 'ecological optics' is a theory of natural vision, and not yet a theory of the cultural variation which I claim pictorial representation introduces into visual perception. Thus, though my view is compatible with his, at the level of animal or naive vision, it diverges from his in the claim that our vision, as *human* vision, goes beyond the framework of ecology, as an account of the evolutionary conditions of the development and functioning of the visual system, to an account of the *post*-biological or *post*-evolutionary – i.e. cultural and historical – conditions of the development and functioning of the human visual system. What I am offering is therefore *not* the experimental evidence on which my proposed theory is based, but rather a theoretical proposal which will *lead* to experimental or empirical research.

Not that there is not a large and growing body of data, which is already relevant to this view. But rather, that the interpretations of this data are under-determined, and the type of research which is presently going on has not yet posed sharply enough the relation between prevailing modes of pictorial representation and prevailing modes of perception. The areas of research are in cross-cultural studies of picture-perception;[12] and in experimental studies of perceptual interpretation of objects and scenes in the three-dimensional world, on the one hand, and of two-dimensional representations or pictures of these objects and scenes on the other.[13] In this article, I can only point to the growing body of such studies, and note how vigorous and lively the present discussion and research activity is.

(2) This brings us to the second question: What does it mean to claim that perception is rule-following rather than law-like in its operation? Most radically, this claim suggests that, since rules are created by and changed by historical and cultural practice, therefore, perception itself can change in its mode, with such variation. In short, human perception is historically conditioned, and not simply biologically formed. How would one show this? It seems to me that the whole weight of the argument, in the philosophy and history of science, for the view that experimental observation is to one or another degree theory-conditioned; that what we see is bound to frameworks of interpretation which predispose us to what there is to be seen – in short, that the general view of the comparative and

historical social psychology and sociology of perceptual belief and per-
ceptual activity, argues for the conclusion that perception is rule-
governed. One may argue, against this, that this holds perhaps for the
higher reaches of experimental observation in the sciences, where scien-
tific theory is involved in defining the entities to be observed, and in
defining the means of experimental observation and measurement them-
selves; but that this is not so, (or nor clearly or necessarily so) about
'ordinary' – that is, pre-theoretical, practical, common-sense – percep-
tion. Yet, there is enough evidence at the allegedly 'lower' reaches of
perception – e.g. experiments in word-recognition, picture or object
identification, physiognomic recognition, etc. – to show that variations in
commonplace belief, expectation, 'set' (*Einstellung*), emotional states,
perceptual hypotheses, – condition and change our perception. Still, one
may argue against this view, that such variation holds at higher or more
complex levels of perception, which involve, e.g. social belief, but not at
the rock-botton perception of such fundamental physical and ecological
variables as size and shape. My suggestion here is that just such funda-
mental perceptual variables are *also* rule-governed. The rules are not as
such rules for perception *per se*, but rather rules for pictorial representa-
tion of what is perceived. Thus, the experimental test of such a suggestion
as I put forth here depends on contrasting perceptual contexts governed
by different rules. This is the point of cross-cultural studies of picture-
perception, as well as of differential tests of subjects who are requested to
represent pictorially what they see (or who are requested to choose
among alternative pictorial representations). On my view, then, alterna-
tive rules or canons of the perspective representation of shape or size,
such as are exemplified, e.g. in pre-Renaissance Western painting, or in
Chinese or Persian painting or in so-called 'primitive' art, are not simply
ways of picturing, but also ways of seeing. What I am suggesting is that
there is a significant connection between pictorial styles and what one
may call 'visual styles,' (with respect to the variables of size and shape at
least); and that just as representation has a (cultural) history, so too does
vision, by virtue of its involvement with the activity of pictorial represen-
tation.

   If this is the case, then we are confronted with (at least) two alterna-
tives: either the history of visual styles, like the history of representational
styles, has no *intrinsic* norm which determines which mode of perception

is more 'correct' or 'truer' than another – (different ways of seeing are
simply historical facts about human perception, and one is not 'truer' than
another, or 'more correct'); or visual styles, and concomitantly represen-
tational styles as well, are to be judged by some standard of adequacy,
such as veridicality or the fidelity of perception (and thus too, the fidelity
of pictorial representations of what we perceive). Let me turn, finally,
then, to a discussion of one aspect of this question: whether, and in what
sense, pictorial representation in accordance with the rules of linear
perspective, is a more correct rendering of what we see than those modes
of pictorial representation which do not follow this rule.

What is at issue in the question whether perspective representation is
'correct,' or more correct, or less mistaken than non-perspective rep-
resentation, is whether there is some test for correctness. It is sometimes
argued that the empirical test of the correctness of a representation, or of
its fidelity, is recognizability. If recognizability of perspectival representa-
tion is decisively greater than that of alternatives, then the canons of
perspective representation are awarded the palm for fidelity. What we
need, then, is a good experimental design, i.e. one which is not vacuous.
For if our experimental subjects are already predisposed to measure
pictorial or representational fidelity by the canons or rules of perspective
representation, then of course, our experiment begs the question, and
decides nothing. Experimental psychologists, therefore, have chosen
their experimental subjects in such a way as to aviod this. As I sum-
marized some of the research, in my earlier paper, subjects of various
sorts, from visually naive to sophisticated, from apes, infants and idiots to
teachers of perspective drawing, were chosen.[14] Goodman notes some
older cross-cultural studies and there have been others.[15]

Let me not insist that the outcome bears out Goodman's view, or my
own, though I think it does. What matters is that the view that recogniza-
bility tests fidelity of representation, is simple-minded and mistaken. For
suppose perspective representations *did* turn out to be the most univer-
sally recognizable, i.e. most successfully identifiable as representations·of
their objects, by some statistically significant measure; and that therefore,
perspective representations were judged to be more 'correct' on these
grounds than some alternative. What would remain at issue is *why* this is
so. The standard view proceeds from the premise that the mirror theory
of vision, and the concomitant theory of perceptual constancy give us a

true account of visual reception and perception, where the theory of vision tells us what is received by the eye, and constancy theory tells us how what is thus received is correctly perceived, by means of a transformation. Therefore, a representation which ostensibly duplicates what the eye receives is 'true' in the sense that it yields an image which corresponds to that produced by the objects or scenes represented, thus preserving fidelity in the representation. Note, however, that such an account of fidelity depends on a theoretical premise, namely, the mirror-theory of vision. And our test of fidelity was to be empirical, i.e. recognizability, or identification of the representation with what is represented. The claim that recognizability depends on fidelity in *this* sense (of perspective representation) is vacuous, however, if it can be shown that recognizability is equally achieved, or better achieved by other means; or if it can be shown that the theory which describes what the eye receives is false, and that therefore the account of fidelity, based on the duplication by representation of the received visual image, is itself false.

But let us grant, for the moment, the assumption of the 'truthfulness' of the perspective representation, in the sense of a faithful reproduction of what the eye receives. If it is the case that some alternative mode of representation yields as great, or greater recognizability, without *this* kind of (perspectival) fidelity, then this fidelity in itself is neither the unique nor the guaranteed criterion of recognizability. Indeed, if what we mean by fidelity, functionally, is recognizability itself, and if degrees of fidelity are correlated with degrees of (experimentally testable) recognition, then a quickly recognizable caricature would be a more faithful representation than a careful photograph which may yet be difficult to identify with its subject. (One may then even go so far as to claim that if recognizability is the *only* criterion of fidelity, then, strictly speaking, labels or titles on paintings would count as means of recognition, though they are in no way representations.) Recognizability is not the burden of a particular rule of representation, then, but may be achieved by alternative rules, or may even be a function of completely extraneous, non-representational factors. But if this is so, recognizability as such is no test of fidelity in the usual sense of resemblance. [16]

Yet, we are hard put to rid ourselves of the nagging sense that perspective representations are after all, more faithful, truer to the way things *really* look, and more recognizable. My argument is not that this is

not so, but that it is so precisely because we have adopted the rule of perspective representation as our norm of fidelity, and not because some independent criterion of fidelity (e.g. recognizability) has dictated that we should adopt this norm rather than some other. In short, it is the choice of a norm of fidelity that affects recognizability. And in human perception, such norms are achieved, and not merely given with our physiology.

To conclude: I have argued that the theory of perceptual constancy is based on a mistake. The mistake is that the theory supposes that what we receive, as visual input, is variable in the way which a given theory of vision describes it. But this theory of vision itself takes the visual world to be a *picture* and proposes that we see it *as* a picture. My counter-argument is that we see the visual world as a picture because we picture it in certain ways. And therefore, what we see *becomes*, in significant part, a function of our modes of picturing. Since these modes change, historically and culturally, so too does our mode of visual perception itself. Seen in this perspective, the theory of perceptual constancy is not simply mistaken, but rather mistaken in its object, i.e. in what it is a theory about. Taken as a theory of our visual perception of the world, it is a mistaken theory, as I tried to show. Taken, however, as a theory of perceiving invariances in pictorial representations which are themselves made in accordance with the rules of linear perspective, the theory not only *may* be correct; it *must* be correct. For the theory itself *defines* and *prescribes* what the variations should be for which it provides the invariances through transformation.

Boston University

NOTES

[1] Euclid's *Optics* (c. 300 B.C.) already gives the essential theorems for these transformations. The interpretation of the eye as camera, with the projection of an inverted image on the retina, by means of convergence of the light rays by a lens, is formulated by Kepler, *Ad Vitellionem Paralipomena* (1604), and Descartes (*Dioptrique*, 1637) among others. The pinhole *camera obscura*, which forms such an inverted image, was already used by, and theoretically understood by Alhazen (c. 965–1039). Kepler's work is a commentary on Vitellio (13th century), whose work is a commentary on Alhazen. For an excellent account of the elements and a brief history of the theory of vision, see M. H. Pirenne, *Optics, Painting and Photography*, Cambridge: Cambridge University Press, 1970, chapters 1–6.

[2] The fundamental error of conceiving of the retinal image as something we 'see,' or which somehow presents itself to our awareness as an internal 'picture' comes from viewing the eye as a camera. As M. H. Pirenne points out, "we do *not* see our retinal image . . . . As LeGrand has said epigrammatically, the eye is the only optical instrument which forms an image which has never been intended to be seen. This is the great difference between the eye and the photographic camera. Failure to realize this lies at the root of many misunderstandings." (in *Optics, Painting, and Photography*, Cambridge: Cambridge University Press, 1970, p. 9). Pirenne points out that Descartes already understood that the retinal image was not itself a 'picture' for us: "Descartes in his *Dioptrique* had already postulated that the retinal excitation pattern was conveyed to the brain, so that a picture ('une peinture') was formed there, bearing a certain resemblance to that formed on the retina and therefore to the external objects. But Descartes himself insisted that it was *not* by virtue of the resemblance of this 'picture' with the objects that we see them *'as if there were again other eyes within our brain with which we could see it'* . . . ." (*loc. cit.*) Newton, however, in his discussion of Axiom VII, in the *Optiks*, seems to propose just such a view of visual perception as a kind of 'seeing' of retinal images, though his language is ambiguous, and he speaks of the retinal images only as the 'causes' of our vision. Thus, he writes, 'In like manner, when a Man views any object PQR, the Light which comes from the several Points of the Object is so refracted by the transparent skins and humours of the Eye, (that is, by the outward coat EFG, called the *Tunica Cornea*, and by the crystalline humour AB which is beyond the Pupil *mk*) as to converge and meet again in so many Points in the bottom of the Eye, and there to paint the Picture of the object upon that skin (called the *Tunica Retina*) with which the bottom of the Eye is covered. For Anatomists, when they have taken off from the bottom of the Eye that outward and most thick Coat called the *Dura Mater*, can then see through the thinner coats, the Pictures of Objects lively painted thereon. And these Pictures, propagated by Motion along the Fibres of the Optick Nerves into the Brain, are the cause of Vision. For accordingly as these Pictures are perfect or imperfect, the Object is seen perfectly or imperfectly." (*Opticks*, New York: Dover, 1952, p. 15.)

[3] See, for example, David H. Hubel, 'The Visual Cortex of the Brain.' *Scientific American*, November 1963; Stephen W. Kuffler, 'Discharge Patterns and Functional Organization of Mammalian Retina', *Journal of Neurophysiology*, January, 1953; David M. Hubel, 'Integrative Processes in Central Visual Pathways of the Cat', *Journal of the Optical Society of America*, January, 1963; D. H. Hubel and T. N. Wiesel, 'Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex', *Journal of Physiology*, January, 1962; Ragnar Granit, 'The Visual Pathway', *The Eye, Volume II: The Visual Process*, Academic Press, 1962; J. Y. Lettvin, H. R. Maturana, W. H. Pitts, and W. S. McCulloch, 'Two Remarks on the Visual System of the Frog', *Sensory Communication*, The M.I.T. Press, 1961; S. W. Ranson (revised by S. L. Clark), *The Anatomy of the Nervous System*, Tenth Edition, W. B. Saunders Company, 1959, pp. 264–266.

[4] See, for example, Pirenne's discussion of the 'corrections' of the 'distortions' of linear perspective by artists, *op. cit.*, pp. 121–135.

[5] M. Wartofsky, 'Pictures, Representation and the Understanding', in R. Rudner and I. Scheffler (eds.), *Logic and Art-Essays in Honor of Nelson Goodman*, Indianapolis and New York: Bobbs-Merrill, 1972, pp. 150–162.

[6] See my discussion of this in M. Wartofsky, 'Perception, Representation and the Forms of Action: Towards an Historical Epistemology', in *Ajatus* Vol. **36**, *Yearbook of the Philosophical Society of Finland: Aesthesis, Essays on the Philosophy of Perception*, 1976, pp. 19–43.

[7] See, for example, T. G. R. Bower, 'Slant Perception and Shape Constancy in Infants', *Science*, Vol. **151**, pp. 832–834; 'The Visual World of Infants', *Scientific American*, Vol. **215**, No. 6, pp. 80-92; and *Development in Infancy*, San Francisco: W. H. Freeman and Company, 1974.

[8] See Robert L. Fantz, 'Pattern Vision in Young Infants', *The Psychological Record*, Vol. **8**, pp. 43-47, and his 'The Origin of Form Perception', *Scientific American*, Vol. **204**, No. 5, pp. 66–72.

[9] See especially the following works of J. J. Gibson: *The Visual World*, Boston: Houghton Mifflin, 1950; *The Senses Considered as Perceptual Systems*, Boston: Houghton Mifflin, 1966; 'The Information available in Pictures', *Leonardo*, 1971, 4, 27–35; and *An Ecological Approach to Visual Perception* (forthcoming).

See also the excellent discussion of Gibson's view in relation to these issues in Margaret Hagen, 'Picture Perception: Toward a Theoretical Model', *Psychological Bulletin*, Vol. **81**, No. 8 (1974), pp. 471–497; and also the Gibsonian approach in John M. Kennedy, *A Psychology of Picture Perception*, San Francisco: Jossey-Bass Publishers, 1974.

[10] This view is developed in an as yet unpublished series of lectures on historical epistemology, available in draft form upon request from the author.

[11] See R. H. Thouless, 'Phenomenal Regression to the Real Object', Parts I–II, *British Journal of Psychology*, Vols. **21-22** (1931); 'Individual Differences in Phenomenal Regression', *British Journal of Psychology*, Vol. **22** (1932); and his later article, 'Perceptual Constancy or Perceptual Compromise', *Australian Journal of Psychology*, Vol. **22** (1972).

[12] The literature here is vast. See the large bibliography in Margaret Hagen and Rebecca K. Jones, 'Cultural Effects on Pictorial Perception: How Many Words is One Picture Really Worth', in Walk and Pick (eds.), *Perception and Experience* (forthcoming). See also Jan B. Deregowski, 'Pictorial Perception and Culture', *Scientific American*, Vol. **227**, (1972), pp. 82–88; and John M. Kennedy, *op cit.*, Chapter 5 ('Picture Perception across Culture and Species'); and M. H. Segall, D. T. Campbell, and M. J. Herskovits, *The Influence of Culture on Visual Perception*, Indianapolis: Bobbs-Merrill, 1966.

[13] This is the subject of present experimental research by Prof. Margaret Hagen of the Department of Psychology at Boston University.

[14] See 'Pictures, Representation and the Understanding', *op. cit.*, pp. 155; and H. Leibowitz, I. Waskow, N. Loeffler, and F. Glaser, 'Intelligence Level as a Variable in the Perception of Shape', *Quarterly Journal of Experimental Psychology*, Vol. **11**, (1959), pp. 108–112.

[15] For Example, M. Herskovitz, *Man and His Works*, New York: Knopf, 1948. Goodman quotes this, in *Languages of Art*, fn. 15, p. 15. See above, fn. 12, for reference to further cross-cultural studies.

[16] See Robert Schwartz, 'Representation and Resemblance', *The Philosophical Forum*, Vol. **V**, No. 4 (1974), pp. 499–511, for an excellent discussion on this point.

ACTION AND PASSION: SPINOZA'S CONSTRUCTION OF A SCIENTIFIC PSYCHOLOGY

[1973]

1. INTRODUCTION

Spinoza's construction of a scientific psychology is one of the most striking historical examples of the heuristic function of metaphysics in the genesis of scientific theory.[1] It is, at the same time, an example of how the requirements of a scientific theory are related to the construction of a metaphysics. That these two propositions are not mutually exclusive, I hope to show in this essay; and that they are both true requires us only to believe that science and metaphysics mutually interact, and help to shape each other, especially in those periods of great discovery and courageous theorizing which mark the youth of a new science.

What is at issue in this essay is Spinoza's claim that a science of human nature is continuous with a science of nature; that human action and passion are as subject to universal laws, and therefore as subject to rational understanding, as is the motion of bodies, in physics, or the relations among points, planes, and solids in geometry. To understand Spinoza's claim, I hope to specify the particular *Problematik* with which he was presented by Cartesian psychology, and to show the forcefulness both of his methodological and psychological insights. But beyond this, I hope to make clear a programmatic point: to show how Spinoza's conception of a science of psychology is related to his metaphysics, and in particular to his ontology of the person, as a natural individual continuous with all of nature, yet distinct as a conscious organism, i.e. as organized matter which thinks and feels, and which acts in order to survive, or to preserve its individuality.

To begin, I will consider the phrase 'scientific psychology', both with respect to the notion of a method or a kind of knowledge, and with respect to the specific characterization of its domain. I will then present the *Problematik* or problem-setting of Spinoza's construction, and then proceed to the reconstruction of Spinoza's psychology as a theoretical system, embedded in and supported by a metaphysics. Here, the notions 'action' and 'passion' will be seen to be theoretically central, as will the

epistemological notions 'adequate' and 'inadequate ideas'. Finally, after a consideration of Spinoza's 'mechanics' of the affects, or emotions, I hope to point to the relevance of Spinoza's program for contemporary psychology.

2. 'SCIENTIFIC PSYCHOLOGY' IN HISTORICAL CONTEXT

The phrase 'scientific psychology' I take here as expressing a historical conception, and not some canonical contemporary 'science'. First, it can hardly be claimed that there is a canonical 'scientific psychology' today. Second, the shifting conception of what is canonically 'scientific', in general, cannot be understood as more than an arbitrary sequence of fads and faiths, unless the genesis of our contemporary conceptions is understood historically. To this end, the theory of science which I want to present, specifically, is that which developed in the crucible of seventeenth-century philosophical debate and scientific discovery, and to which Spinoza's most crucial contribution was his psychology. How this theory of science comports with our own theories is a separate and difficult question, about which I will say very little here. A general methodological approach to the question of what constitutes a Spinozist 'scientific psychology' begins with the question, *first, of what constitutes scientific knowledge* for Spinoza; and *second, of what constitutes the specific domain, or the specific object of this knowledge.* To the first question, the answer is Spinoza's own characterization of scientific knowledge as the 'second kind of knowledge', namely, that which realizes the lawful connections between phenomena, and comprehends the causal or determinate manner of these connections. In practice, such comprehension or understanding is exhibited in the formal statement of these connections in a deductive system, such that the implicit conditions of the famous theorem ($E.$, II, vii: "The order and connection of thoughts is the same as the order and connection of things") are fulfilled. Scientific knowledge, therefore, is the affirmation of true thoughts, i.e. of those which are in agreement with the order and connection of things; nor is this simply an ordering sequence; the 'connection' is causal, determinate, and in Spinoza's special sense, necessary; namely, it follows from the nature of Substance itself. The way things are, they are necessarily; and so, a science, as the systematic idea of nature, is a knowledge of this necessity *as* necessary, and not simply under the form of contingency. (E., II, xliv, & Cor. 2. Dem.) The limits of a science are therefore the limits of the clear

and distinct, i.e. the 'adequate' ideas we can have, and are therefore related to the conditions and limits of our knowledge. Spinoza's conception of scientific knowledge is therefore closely related to his psychology, insofar as the psychological theory gives us in turn an account of the genesis and constraints on our knowledge.

This brings up, quite specifically, the related question of the domain or the object of a science; since what can be known of a particular domain or object depends on the way in which knowledge of it is possible. In the case of psychology, the domain or object of knowledge is man's conscious action, i.e. human activity itself, insofar as it is an object of consciousness, and can come to be known in accordance with laws and principles. More precisely, for Spinoza, psychology includes the mind's knowledge of its own body, insofar as this body undergoes changes in its 'power of acting'; and the mind's knowledge of its own affections, insofar as these are conceived as affections of the mind itself. The first of these constitutes the domain of imagination, perception, and the emotions; the second, of thought proper. Spinoza offers us both a theory about how such knowledge (of the emotions and perception and of thought) is possible – this is the epistemological groundwork of his psychology – and also, a psychological theory of emotion and thought – that is, an account of the *laws* of emotion and of thought, as natural phenomena, and therefore as part of a general science of nature; in particular, that part of it which is the science of man.

3. SPINOZA'S PROBLEMATIK

The *Problematik* of Spinoza's psychology is posed by Cartesian psychology and its metaphysical framework: namely, by mind-body dualism. Insofar as this concerns Descartes's psychology, it can be summarized by two basic ideas: first, that the science of bodies, in their motion and interaction, is a mechanical physics, whose ontology is that of inert matter, whose principle of motion lies outside itself. Animal bodies are subject to the same mechanism as the rest of physical nature; and insofar as animal bodies are affected by motion or change, this motion or change can be understood as a composition of motions of the parts of animals, and of the interaction of these bodily parts with external bodies. Second, however, insofar as these motions or changes affect conscious beings, the mode of this affect is by mechanical and causal interaction of bodily affections with the soul, by the mediation of

the pineal gland. But as opposed to the inert and extended property of matter, in which only efficient causes operate, and in which all motion is that of moved movers, the soul or thinking substance has its principle of motion in itself, and is, as soul, fully self-determined, and in this sense has agency, will, and freedom.

In Descartes's psychology, therefore, the automatism of body is sharply contrasted with the autonomy of the soul; the science of the one is mechanics; of the other, autonomous reason. The psychology of Descartes poses a double problem: first, insofar as it is psychology proper – i.e. insofar as it deals with *psyche* or *soul* in itself, it deals with the interaction of the bodily affections with the soul only at its margins, if at all. Second, if psychology is to deal with this interaction, it is faced with the inordinate metaphysical difficulties of the interface between mutually exclusive ontologies and mutually exclusive methodologies: how can the mechanical interact with the autonomous, the extended with the nonextended, the finite with the infinite, the determined with the free?

This excursus sets the problem of Spinoza's psychology sharply, and also permits a characterization of his solution to it: not only the bodily affections but those of the mind as well are to be included in the science of mechanism. The continuity of sensory perception, emotion, and thought is to be reaffirmed. But such a reaffirmation cannot be achieved simply by methodological fiat; rather, the metaphysics of Descartes has to be fundamentally revised, and the ontological dualism overcome, so that a methodological monism can be asserted. If no domain is to be immune to the mathematical method, and to explanation in terms of efficient causes which this method offers, then mind, no less than body, must be adequately conceived as determined to its activity by causes. But if mind is to be conceived, as it must be, as an activity whose principle of motion and change is in itself, then so too must body. The dualism of inert substance on the one hand and self-active substance on the other, must be exchanged for a monism of self-active substance, whose modes of activity are differentiated but whose principle of activity is not. Such a self-active principle must be such that its existence and its activity are one and the same; it must be therefore *causa sui* in the special sense of an *active causa sui*; its being must be identical with its activity; and it must, at the same time, be one, and yet be self-differentiated; its 'parts' must therefore be conceived not as simple mechanical divisions, but as partial expressions of the whole.[2] Spinoza therefore sets himself the task of

constructing a psychology which, unlike Descartes's, includes the soul within the context of a science of mechanics and yet retains its self-activity. Both mind and matter therefore need to be recast for the requirements of this monism; and the continuity of the bodily affections with the mind's activity has to be systematically worked out. So, too, the consequences and limits of mechanism itself are tested, and the inadequacy of mechanism begins to be shown. In effect, then, Spinoza's extension of mechanism begins to transcend this mechanism itself and therefore to strain the very conception of the prevailing scientific methodology.[3]

I am arguing that Spinoza's metaphysics was not a philosophical or theological exercise in its own right; but rather, that its motivation was the problem of a consistent scientific methodology which would include the domain of psychology; that the metaphysics was in the service of this aim. It may be argued that this view fails to grasp the essentially ethical motive in Spinoza's thought. But the continuity between the ethical and the scientific in Spinoza, as in Aristotle, lies in the conception of the ethical as that activity which is in accordance with the nature of man, and the discovery of this nature and of its proper activity is the task of rational science in the service of man's well-being. Moreover, Spinoza's conviction is that a wrongly conceived morality, based on superstition, fantasy, wishful thinking, is the product of human ignorance, and only the critique of this superstition frees man for his proper activity, in which alone his happiness and well-being reside.

That Spinoza fails in his task, that his monism suffers from inconsistencies and obscurities, that his psychology itself founders on the discontinuities of body and mind, as he himself conceives them, is the burden of much of Spinoza criticism. Yet the towering attempt at a system, and the power of its monistic imperative still provide, to my mind, one of the most viable heuristic guidelines in the formulation of a contemporary science of psychology.

4. THE THEORETICAL CONSTRUCTION, I: BODIES MINDS, AND IDEAS

I will treat Spinoza's psychology principally from the point of view of its systematic construction. The theory of action and passion in Spinoza is in effect his theory of the nature of the person and the primary theoretical construct of his psychology. To Descartes's divided being,

constituted by body and mind, Spinoza counterposes an active, integral organism, which is essentially body, constituted of parts. This composite body is identical with its activity, or more precisely, with its power to act. Moreover, this composite is not merely an aggregation of parts, but an individual thing. (See Spinoza's argument on composition, *E.*, II, xiii, Def. following Axiom 1, and Lemmas 5, 7; Schol.; Post. 1.) Identical with this composite or organic individual is the human mind, which is, in Spinoza's phrase, the idea of this body. It is one and the same thing, as body, conceived under the attribute of extension, and as mind, conceived under the attribute of thought or consciousness. The person, or the human individual, is therefore a body-mind, that is to say, a determinate mode of substance, conceivable under both attributes, but self-identical, under both. As a determinate mode of substance – that is to say, an individual thing – it is by definition finite. And here, the crucial nature of finite existence, or of the modes, in Spinoza, is the metaphysical clue to his psychology. "Every determination," says Spinoza, "is a negation." That is to say, determinate being is 'caused' by something external to it, which defines its limits, makes it the particular individual it is, and characterizes its powers to act. No determinate being can act with infinite power, precisely because as determinate, its power (and synonymously, its existence) is bounded by its (necessary) relations with all other determinate things. As determinate, it is essentially *interactive*, or is both caused *by* and is the cause *of* other things. Nothing is undetermined, for then it would not exist, or be an individual. But everything (short of Substance itself) is both *determined* and *determining*; passive, insofar as it is determined by something other than itself; active, insofar as it determines some other, or itself. Only the whole system or universe of such interactions has no delimitation, since by definition it is the whole, and there is nothing external to it. This infinite being, or substance, therefore, doesn't 'act' *on* anything, but is identical with all of the internal activities of its modes; not only is it infinite activity, in this metaphysical sense, but it is infinitely self-differentiated activity, since the chain of causes, according to Spinoza, is infinite, without beginning or end. (See, e.g. *E.*, I, xxviii & II, xiii, Lemma 3, for Spinoza's rejection of infinite regress arguments.)

In what sense, then, can a body or an individual be active, if its activity is fully determinate – i.e. if it is caused to act by other individuals with which it interacts? Here the very language and conception of mechanism are strained beyond limit. For, being the very individual it is, it is not

simply constituted as the set of determinate relations it has with
everything else in the universe – though it is also that; rather, the quality
and character of its relations, or its actual and potential interactions with
other things are also determined by its own composition, i.e. by the
simple or composite bodies which constitute it. The dialectical dilemma
of a thing *constituted* by no more than its relations was resolved by
Leibniz in his postulation of the monads as dimensionless, mathematical
points. This relational *existence*, however, sacrifices the materiality of
the basic individuals of the system. By contrast, the simplest bodies, for
Spinoza, are extended, and therefore, so too are the composites. They
are material bodies which, by virtue of their power to affect other bodies,
are active, and by virtue of their capacity to receive the actions of other
bodies, are passive. Insofar as such a body acts, therefore, the cause of
its activity is in itself, and can be conceived clearly and distinctly by the
mind, as following from the nature of the body. In Spinoza's language,
"we act when anything is done, either within us or without us, of which
we are the adequate cause, that is to say . . . when from our nature
anything follows, either within us or without us, which by that nature
alone can be clearly and distinctly understood" (*E.*, III, Def. 2). But in
fact, Spinoza goes on to say that though the mind *is* the idea of the body,
and is thus necessarily aware of everything that happens in its object, it
comes to know its object, i.e. its body, only by virtue of its awareness of
changes in the body, i.e. of modifications or affections of the body. On
the one hand, what we can know clearly and distinctly of the body is
what Spinoza calls common notions – i.e. those ideas which all men have
in common, concerning body, and what is common both to our own
bodies and to external bodies. But insofar as we have an adequate idea of
the body, this idea must be the idea of the body as an adequate cause,
that is to say, of the body acting in accordance with its nature. And thus,
we cannot have an adequate idea of the body insofar as it is affected by
other bodies external to it. But since the body, as determinate, is what it
is in its interaction with other bodies, and its determination or
modification by them, we cannot have an adequate idea of the body,
short of having an adequate idea of the whole system of interactions.
Thus the finitude of the body is at the same time the finitude of the
human mind. Its knowledge of itself, as an individual, is forever limited
by its partiality with respect to the total scheme of things. But this
knowledge is not as such false, thereby; it is only the *occasion* of falsity,
or the possibility of error.

The upshot is that the body is the adequate cause of its actions only as it is seen as substance acting, in one of its modifications, i.e. only as the whole system of interactions is expressed in it, or only as the mind can come to conceive it under the form of necessity or under the form of eternity. Insofar as the mind is the idea of the body, and the body is a determinate body or an individual in a system of such individuals, the mind is, as is the body, a finite mode, a part of this sytem. What it can come to know is the systematic interrelation of all other bodies to its own, but only insofar as the mind's own body is affected by these interactions. Therefore, it can never come to know external bodies (or causes) in themselves, but only by the effects they have on the mind's body. This knowledge, Spinoza says, is knowledge which the mind *necessarily* has, since it *is* the idea of the body, and is not a separate or derived reflection *of* the body. In effect, the identity and being of the mind is the consciousness of bodily affections; or better, *is* these bodily affections conceived under the attribute of thought. Thus, the mind cannot *but* have ideas, and thus has them necessarily (*E.*, II, xii). But from this necessary knowledge of one's own bodily affections, it does not follow, says Spinoza, that we have *adequate* ideas of the external bodies insofar as these are the sources of these affections. And insofar as the body is affected from without, the knowledge of the bodily affection itself is inadequate since it is not itself the cause of its affections, but only their partial cause.

The individual human body as a finite mode of substance is part of a system of such bodies; and thereby, interaction and interdependence are the very modes of existence of such bodies [Postulate 4, following *E*. II, xiii]. Yet, Spinoza's notion of such bodies as composites of bodies, and as composites of composites, constituting at each level of organization a unity or an individual, permits him to ascend from 'simple bodies' to the one 'individual' comprised of the system as a whole, and therefore, to the idea of that one individual as the idea of the whole, whose awareness of that whole (its 'body') is an awareness of an infinitely differentiated unity; moreover, of this unity as containing all of its differentiation as *self*-differentiation, and therefore as its own activity (since nothing is external to it), of which it necessarily has an adequate idea. At this limit, as Spinoza expresses it, mind achieves "the intellectual love of God," or contemplates itself accompanied by the idea of God as cause. It achieves, in effect, knowledge of individual objects as they are in themselves, by virtue of this knowledge of the 'third kind', or intuition (*E.*, II, xlvii and Schol.).

Short of this state of blessedness, with which Spinoza concludes the
Ethics, there is the finitude of our existence, as its necessary condition.
Spinoza relies here on common sense as much as on metaphysics; or
rather, his metaphysics simply states, in systematic and abstract fashion,
the requirements of common sense: as individuals, we are dependent
upon and interdependent with other individuals. As human beings, our
existence *is* in this interaction, both as bodily beings and as conscious
beings. The consciousness of our interdependence is not simply a con-
dition of the finitude of our minds, but equally (and identically) the
condition of our bodily existence. But precisely because of this necessary
condition of dependence upon others, and because we can have adequate
ideas only insofar as these are of our own activity, or what follows from
our own natures, then insofar as we interact with external individuals,
our ideas are inadequate or confused; and we are, by nature (the nature
of our finitude or dependency), condemned to inadequate ideas of this
interaction. We are therefore the subjects of passions as well as actions,
and therefore, insofar as we are human, we suffer. For the passions
according to Spinoza (in common with Descartes, and a long tradition)
are those affections of the body of which we are not ourselves the cause,
except partially; or are those changes in our power of acting which are, in
part at least, impressed upon us from without. Since by definition such
affections are known only inadequately, then insofar as we have an
inadequate idea of anything, we suffer or are subject to the passions, and
in proportion to the number of inadequate ideas we have.

Thus, Spinoza's theoretical construction of a science of psychology
begins with his notion of body, of the action of bodies and of affections
of the body. A body, however simple, is extended; and moreover, is
identical with its activity, or power of acting. As an individual, short of
being the composite and unique individual which is the universe itself
(conceived under the attribute of extension), a body is part of a system of
bodies; and is itself constituted as a system of bodies, which are its parts,
insofar as it is not a simple body. Human beings, since they are neither
atoms nor simple bodies, on the one hand, nor Substance itself, nor God,
on the other, are composites with bodily parts, and are, as individuals,
part of larger composites or systems of individuals. It is this position
midway in the scale of nature – neither its ultimate constituents nor the
whole – which constitutes the ontological character of the human
individual, or person; and which provides the theoretical framework for
an account of him as a conscious being, whose agency is identical with his

body's power of acting. This power of acting, or self-activity is, in
Spinoza's terms, the "perfection," the "reality" or the "existence" of
any individual. The more modes of acting an individual has, the more
"reality" or "perfection" such an individual has. And this power of
acting is therefore dependent upon the kinds and modes of interaction
available to that individual. Thus, the dependency on other bodies, in a
strange and dialectical sense, is the very condition of a body's activity,
since its power to act is its power to affect other bodies; as, in turn, the
power to act of these other bodies is their power to act on this (my) body.
The fundamental mode of the existence of human bodies, as individuals,
is therefore a relational mode, or one of interaction.[4]

   Insofar as composite bodies are acted upon by other bodies, they
undergo modifications or affections. Spinoza treats this in a thoroughly
mechanistic manner, in that such affections are literally changes or
"traces" impressed upon the body, or its parts. When such changes or
modifications disrupt the "proportion of motion to rest," which is the
equilibrium-condition for the continued existence of an individual – i.e.
when the composite relation of motion and rest of the parts of the body is
disturbed – then the individual, as *that* composite, may be destroyed. But
short of this, the impressions or traces left on a composite individual by
the actions upon it of other external bodies are affections which are
necessarily "known" in the body, as the idea of that body, or the mind.
The mind's awareness of such affections (or "the idea of these bodily
affections") is, according to Spinoza, the very essence of what it is to be
a mind; and in this sense, the mind is the "idea of the body." But
Spinoza's theory of how these affections come to be known depends on a
mechanism of bodily affections which produces "images" – on
Spinoza's account, a kind of physiological "echo" effect, whereby an
impression made by an external body, striking on the "fluid parts" of
the body, causes a deflection in the plane of the "softer parts," and
thereafter the fluid parts "by their own spontaneous motion" are
reflected in the same way (i.e. by an equal angle of reflection) by this
changed plane, as they were upon their original incidence. By this "angle
of incidence equals angle of reflection" mechanism, together with a
notion of a reverberation or continuing "spontaneous motion" of the
fluid parts, Spinoza constructs a theory of reflection, or image-
formation, whereby the affection of a body by an incident body may be
continued when the incident body is no longer acting or present. In short,
Spinoza proposes both a "trace" theory of images, and a mechanism for

memory – i.e. of the persistence of images when the actions or bodies originally producing them are no longer present. A bodily affection is therefore, literally, a change in the bodily constitution; and the awareness of this change is the imagination, i.e. the mind's idea of this change formed necessarily. What is important in Spinoza's theory is that the mind is not *caused* to become aware of an image, by the change in the body; this latter is the Cartesian model of mind-body interaction. Rather, for Spinoza, the mind's idea of this bodily affection is just this very affection itself, conceived under the attribute of thought. It is, therefore, identical, as idea, with the bodily affection. Therefore, too, images are veridical, for Spinoza; they cannot be mistaken because, in effect, they are the ideas of bodily affections, and are so necessarily; i.e. they cannot be otherwise than as they are. Spinoza says, "these imaginations of the mind, regarded by themselves, contain no error, and . . . the mind is not in error because it imagines, but only insofar as it is wanting in an idea which excludes the existence of those things which it imagines at present" (*E.*, II, xvii, S.). Error, for Spinoza, is negative, a deficiency, rather than a positive activity of the mind.[5] Thus the imagination, in presenting things which are absent as if they were present, is the occasion for, but not the cause of, error insofar as the mind is wanting in an idea which "excludes" (in Spinoza's terms) the idea of the presence of the imagined thing.

In connection with his theory of imagination, and the derived theory of memory, Spinoza develops an associationist mechanism, whereby the simultaneous action of two or more bodies on the affected body generates an association in the mind of the image of one with the other. In the proposition (*E.*, II, xviii) where he proposes this, he talks only of simultaneity; in the Scholium to that proposition, he talks also of associative memory in terms of the concatenation of order of the bodily affections. Here (in defining memory as "a certain concatenation of ideas, involving the nature of things which are outside the human body, a concatenation which corresponds in the mind to the order and concatenation of the affections of the human body"), he adds the idea of a sequence-association as well as a simultaneity-association.

The importance of this theory of image-formation and of imagination is that, in it, the image always carries with it the idea of an external body; and therefore, though the mind knows its own body only in the ideas it forms of the bodily affections, it has also the idea of external bodies as part of its idea of these affections, insofar as these affections are images.

Moreover, the idea of the body is thus mixed with the idea of external bodies, as affecting it. And thereby, the mind knows its own body not as the adequate cause of these affections, but only as the partial cause. It therefore has, in Spinoza's terms, a confused knowledge of the bodily affections, since they do not follow simply from the nature of the body itself, or from its activity alone, but also from the action upon it of other bodies. Insofar as these actions upon it change or affect the body's power of acting, they also affect the mind's power of acting – i.e. they are the condition of inadequate ideas, or the limits upon the mind's clear and distinct ideas. For, Spinoza argues, "The mind doesn't know itself except insofar as it perceives the ideas of the affections of the body" (*E.*, II, xxii). In this second-order knowledge, whereby the mind knows not only the bodily affections, as ideas it has of them, but also knows the *ideas of these ideas* of the bodily affections, the mind in effect knows itself. But as these ideas of ideas are themselves tainted with the inadequacy of the bodily affections, since they are not causes of themselves, the mind necessarily has inadequate or confused knowledge of itself as well (*E.*, II, xxix). In this condition of what Spinoza calls "external perception" he locates the Passions, and relates them to inadequate ideas, i.e. to confused knowledge.

One may ask why Spinoza essays such an elaborate theory of bodies and bodily affections as a condition for a psychology, and especially for a psychology of the affects or the emotions. It is precisely because his theory of the affects is intended to be continuous with his natural philosophy, his psychology continuous with his physics. The groundwork of Part Two of the *Ethics*, on bodies and their interaction, and on the identity of mind with the body's power of acting and its affections, is fully utilized in Part Three, "On the Origin and Nature of the Affects."

Several things may be noted here: first, that the metaphysical notions of adequate and inadequate causes has been "reduced" to a mechanics of bodily interactions. The reduction, however, has introduced the metaphysical notion of self-activity (the characterization of substance, as *Natura Naturans*) at the level of extended bodies, i.e. to individuated matter; and further, has defined body in terms of this self-activity, and of the limits imposed upon it, by the self-activity of other bodies. Second, by the mind-body identity, the adequate and inadequate causes have their counterpart in adequate and inadequate ideas. Further, the notions of action and passion are dependent upon the distinction between adequate and inadequate ideas (and concomitantly, adequate

and inadequate causes). Finally, the whole structure of Spinoza's psychology rests on the analysis of action and passion, insofar as these are the systematic and theoretical concepts in terms of which the human beings' "power of acting" is defined. And this "power of acting" is identical with human existence and its psychological characterization – the human being *is* his activity, insofar as he is both agent and patient, in his interaction with other human beings, and with the things which are necessary for his existence. This activity is, according to Spinoza, self-preservative in its nature – it conduces to continue the human being in existence. Interestingly, Spinoza's rejection of all final causes yet preserves *this* one as the essential one: survival is the *conatus*, end and mode of human existence. What conduces to it is good; what affects it adversely is evil; but 'good' or 'evil' only insofar as the mind characterizes what it desires, or what gives it joy or causes sorrow. It is, for Spinoza, the primary affects of *desire, joy,* and *sorrow* which characterize the *conatus*, or the species of self-activity of the human being; and these, insofar as they enhance or increase the body's power of acting, or diminish it, lend it more 'reality' or 'perfection' or 'existence,' or less.

   Spinoza will derive his ethical consequences from this *conatus*, just as Aristotle did, in the best naturalistic fashion. Happiness, after all, is that activity which is in accordance with man's nature. But man's 'nature' is to survive, to persevere in existence; moreover, to 'increase' his existence, or enhance it, by increasing the degree or amount of self-activity, or of action, and by diminishing the constraints on this self-activity, i.e. by controlling the passions. The remarkable thing, apart from questions of systematic success or failure, is the absolute *chutzpah*, the brashness of Spinoza's program: from the motions and interactions of bodies, to the psychology of the affects, to the therapeutic theory by which actions are enhanced and passions regulated, to the vision of blessedness and beatific virtue, all in one continuous sweep, deriving each later or higher stage from the one previous.

5. THE THEORETICAL CONSTRUCTION, II: ACTIONS AND PASSIONS

Spinoza begins his construction with a methodological claim: namely, that the affects should be treated in the same way, by a rational science, as any other natural phenomena; and this, for the reason that the human

affects or emotions *are* natural phenomena, continuous with all of nature. In his celebrated phrase, man is not a "kingdom within a kingdom", but rather is wholly within the *one* kingdom, nature. He is not a disturbance or a break in the continuity and unity of nature; and therefore, there is no realm in which man has absolute dominion, or 'freedom of will', in violation of the universal determinism of the natural world. Neither does man's capacity for folly and vice fall outside the dignity of a natural science:

Nothing happens in nature which can be attributed to any vice of nature, for she is always the same and everywhere one. Her virtue is the same, and her power of acting; that is to say, her laws and rules, according to which all things are and are changed from form to form, are everywhere and always the same; so that there must also be one and the same method of understanding the nature of all things whatsoever, that is to say, by the universal laws and rules of nature. The affects, therefore, of hatred, anger, envy, considered in themselves, follow from the same necessity and virtue of nature as other individual things; they have therefore certain causes through which they are to be understood, and certain properties which are worthy of being known as the properties of any other thing in the contemplation alone of which we delight. I shall therefore pursue the same method in considering the nature and strength of the affects and the power of the mind over them which I pursued in our previous discussion of God and the mind, and I shall consider human actions and appetites just as if I were considering lines, planes or bodies. (*E.*, III, 'On the Origin and Nature of the Affects.')

Yet, it would be a mistake to take Spinoza's methodological reduction for an ontological reduction. Just as bodies are not lines, or planes, so too affects are not geometrical entities. The 'universal laws and rules of nature' are not the laws of geometry; but the laws of geometry, of physics, and of psychology are universal, insofar as a common intelligibility underlies our understanding of them. This understanding has its most elaborated form in mathematics, in the method of demonstration from principles. It is this method, rather than the specific content of the mathematical principles, which Spinoza proposes here. Spinoza proposes, therefore, a unified science which would include previously separated domains. Analogously, the science of mechanics, in the seventeenth century, achieved a unification of the two discrete domains of Aristotelian physics, thereby accounting for both terrestrial and celestial phenomena by a unified set of principles and laws. But these unified principles bore only on *bodies* in motion or at rest. The realm of human action, insofar as it was conceived of as agential, was absolutely sundered from this mechanics. It is this breach between the domain of

nature and the domain of mind, or human action, which Spinoza's monism is intended to overcome.

This program of unification is the methodological motive for proceeding 'geometrically', i.e. by a deductive construction, in his theory of affects. Thus, he beings with the definitions of *adequate* and *inadequate causes, actions* and *passions,* and *affects.* We have seen that *adequate cause, adequate idea,* and *action* are parallel constructions, as are *inadequate cause, inadequate idea,* and *passion.* Spinoza defines *adequate cause* in terms of the concepts *effect* and *clear and distinct idea:* Thus, if an effect can be clearly and distinctly perceived, or understood by means of the cause, then the cause is adequate. Now adequate causes are not simply conceptual entities; that of which I can form a clear and distinct idea is actually existing (since clear and distinct ideas do not admit of falsity, or incompleteness). So the adequacy of the cause is to be understood ontologically, as entirely constitutive, by its action, of the effect. Partial constitution of the effect therefore yields inadequate causes, i.e. those whose effects cannot be understood by means of these causes alone. The notion 'cause' therefore connotes efficacy in producing an effect. Or, it connotes an activity which is in itself either adequate or inadequate to produce or constitute an effect. Insofar as it is adequate, Spinoza calls it an action, and we are said to act "when anything is done, either within us or without us, of which we are the adequate cause''; and we are said to suffer "when anything is done within us, or when anything follows from our nature, of which we are not the cause excepting partially'' (*E.,* III, Def. 2). What is "done'', either within us or without us, insofar as it changes our power of acting, either increasing or diminishing it, is, in Spinoza's term, an *affect.* He gives the definition both in terms of modifications or affections of the body, and of the idea of these affections in the mind. In short, *affects* are changes in the person's life activity, enhancing or hindering it. They are never neutral, but entail some dynamic alteration of the vitality, the *conatus* of the person. If we identify the affects with the emotions, then it is clear that Spinoza holds that the life activity of a human being is constituted of a composition of such emotions. They are nothing but the variations in intensity and power of this life-force itself. Thus, we have a field theory of the emotions, in that the emotions are nothing but the dynamics of the life-force, *its* growth and diminution, as such. This stretches the concept of emotion beyond our present ordinary usage, however flexible this may be. For it says, in effect, that the principal

characteristic of the person acting is his *power* of acting. His very existence *is* this power, both as it pertains to the body's activity and the mind's. And therefore, the split between the cognitive and the emotive or affective, or between faculties of thought and feeling, or more sharply, between thought and action, is systematically denied. Thought, as we shall see, is involved itself as a mode of this affective activity, or is this very activity itself insofar as it is reflected upon, or becomes self-reflective. In short, for Spinoza, thought *is* action, under the form of reflection; or is self-conscious action. The imagination operates as a mechanism whereby the affects are related to their objects, or to their imagined objects. Will is seen not as a separate faculty or agency, but only as the mind's affirmation or denial of truth and falsity. As such, the term *will* is an abstract characterization of the concrete and particular acts of volition, which are expressions of appetite or desire, when these are conceived under the attribute of thought alone. In this sense, Spinoza says will and intellect are the same, or a volition and an idea are the same (*E.*, II, xlix). Spinoza's conception of the person as a *conatus*, a life-activity striving to preserve itself in existence, places all of the usual psychological categories within this context, and sees them ultimately as the changes in the power of acting of this life-force, or as the conditions under which such changes take place.

From this glorious monism to the differential characterization of the affects, Spinoza proceeds proposition by proposition. But the details are subordinate to the systematic construction. Action and Passion are the crescendo and diminuendo marks, the dynamics of a life-activity. Ultimately, by virtue of the human body's dependency on what lies beyond it, and of man's consciousness of this dependency, man's actions are surpassed by the passions; he can never be the adequate cause of all the effects requisite for a life. Therefore, he can never be the master of himself, by an act of free-thought, or by sheer self-discipline. In Spinoza's terms (*E.*, IV, ii – iv), "We suffer insofar as we are a part of nature, which part cannot be conceived by itself, nor without the other parts," and "The force by which man perseveres in existence is limited, and infinitely surpassed by the power of external causes"; and further yet, "It is impossible that a man should not be a part of nature, and that he should suffer no changes but those which can be understood through his own nature alone, and of which he is the adequate cause." In short, man is bound by his finitude to the overwhelming power of the passions, and thus to inadequate and confused ideas.

In this darkest and most pessimistic portion of the *Ethics*, Spinoza proceeds again from his definitions and principles to an elaboration of the actual affects themselves, giving us the balance between those which are actions and those which are passions, and arriving at this by a calculus of the affects, or what we may characterize as a mechanics of the affects. Here, as in his discussion of the action and interaction of bodies, the model is one of the composition of forces or motions, the primary 'motions' or affects being three in number.

6. THE THEORETICAL CONSTRUCTION III: THE AFFECTS

Spinoza, unlike Descartes and other previous theorists of the affects, fixes on three 'primary affects.' His is therefore the most economical construction, in terms of 'primitives'. He characterizes these as Desire (*Cupiditas*), Joy or pleasure (*Laetitia*), and Sorrow or pain (*Tristitia*). Desire is the *conatus* itself; namely the mind's consciousness of its own striving or effort to persevere in its being, to survive; but clearly, not simply a consciousness, but the striving itself *and* its consciousness. Insofar as this striving is related to the mind itself, Spinoza calls it *Will* (effectively, Will to Live); insofar as it is related to the body, it is called *Appetite*; but he says that these are the same, distinguished only in that desire is conscious appetite. Furthermore, he says that this Desire is the very essence of man "insofar as it is conceived as determined to any action by any one of his affections"; and further, "By the word 'desire' . . . I understand all the efforts, impulses, appetites, and volitions of a man which vary according to his changing disposition, and not unfrequently are so opposed to one another that he is drawn hither and thither, and knows not whither he ought to turn." (*E.,* III, The Affects, Def. 1 and Explanation). This first of the primary affects presents a special case, in view of our previous characterization of the affects as the dynamics of the life-force, or of the *conatus* itself. Some commentators[6] see in this an inconsistency in Spinoza. For either desire is the essence of man, or it is an affect, i.e. a modification of this essence. But it seems clear that the person's life-activity is not, itself, some abstract universal 'force', or a merely formal essence, for Spinoza, but rather is the concretely constituted and alterable energy or activity of his existence. Thus, when Spinoza says that desire varies according to man's changing "disposition," he is saying no more than that man's power of acting changes, in accordance with the particular and concrete constitution of a

man at a given time, or through some time. For Spinoza, dispositions are structures, and structures dispositions. To be constituted a certain way – i.e. for the composite human being to exist as that *particular* composite individual – is to be able to act a certain way. It is not the case that a given structure or composition 'exists', and is only then affected by this or that desire, but rather that desire is the temporally indefinite (though finite) mode of activity which expresses, or is identical, as activity, with a given structure. As man's very essence, it is identical with his existence. One may say, with Spinoza, "No desire, no life." Now this *conatus*, which is the ongoing life-force or effort of a given individual which perseveres in existence, may be abstractly characterized as coextensive or identical with the existence or life of that individual. It is not thereby an abstract essence, but rather *is* the form and modification of this striving itself, in the course of a life. This is an exceptionally dynamic or energistic view of the person, as identical with his life activity; and of his modifications (bodily and mental) as themselves constituting this activity; and therefore of the person's constitution itself as a perduring unity through such changes. But this is, I think, precisely what Spinoza seeks, in breaking away from the mechanism of inert bodies, and the autonomism of a simple, undifferentiated mind, in which alone self-motion exhibits itself.

The relation of Desire to the other two primary affects, Joy and Sorrow, is also systematically complex. Spinoza's goal is not simply to give an account of Joy and Sorrow (or pleasure and pain), but rather to fix them, as affects, within the framework of a composite body (and its mind), acting and undergoing. Thus, in an echo of the Aristotelian account of pleasure and pain,[7] Spinoza sees Joy as "man's passage from a less to a greater perfection", and Sorrow as "man's passage from a greater to a less perfection". Since by perfection Spinoza means the same as reality or existence, or a man's power of acting, Joy is an increase in this power, Sorrow its diminution. Nor are these metaphors, for Spinoza, but literally the characterizations of the degree of autonomy a man has, i.e. the extent to which his actions proceed from his nature, or the extent to which he is the adequate cause of what he does, or of what happens to him. Thus, Desire and Joy, as primary affects, are capable of being actions, or active affects. But to the extent that Desire, Joy, and Sorrow are related to external objects, our knowledge of them is through the imagination, as we have seen; and to the extent of our dependency, or the dependency of our ideas of these affects on what we know only inadequately, all such affects are passions and not actions.

Spinoza seems to be saying that even when our power of acting is increased, if this increase depends on what is external to us, or what doesn't follow from our nature alone – namely, on what is not entirely under the control of our own natures, or what derives from bodily affections caused in us, even partially, by external bodies – then the affects are in bondage to these externals. For mind, this means that we cannot have adequate but only confused ideas of what satisfies our desire, or causes us joy (or sorrow). This follows from the fact that the mind cannot have an independent and adequate idea of the external bodies with which the passive affects are concerned, but only an idea of the affections of its own body produced by the external body, and the imagination's positing of an actually existent and present body as the cause of these affections. Yet, for Spinoza, this dependency is not simply negative – it is so only in the case of sorrow or pain – but may be positive as well, in that even such inadequate causes may enhance or increase our power of acting, for else the concept of joy, or of its derivative affects (love, hope, confidence, self-exaltation, etc.) would be meaningless in the context of passive emotions.

There is a dialectical difficulty here, in that our power of acting is presumably that which conduces to our actions, not our passions, and is therefore enhanced by the clear and distinct, or adequate ideas we have of the causes of our affects. But insofar as these affects are passive, i.e. have at least part of their causes in external bodies of which we cannot have adequate ideas, they are limitations upon our activity, and are, as Spinoza says, passions. But how can passions increase our power of acting? Obviously, they cannot, on Spinoza's theory; yet the affect Joy and its derivatives plainly *are*, by definition, increases in our power of acting. Here Spinoza is ingenious, though I think not successful. The sense in which most passions *increase* our power of acting is the sense in which they are counterposed to other passions which *decrease* our power of acting; and the resultant effect of this composition of forces is a *lesser decrease* (by virtue of that, a relative increment) in our power of acting. Spinoza nowhere puts it in this way; yet, in his theory, an affect can be destroyed or delimited only by a stronger and opposed affect (*E.*, IV, vii). True ideas, he says, cannot alter an affect, insofar as they are true; but only insofar as these ideas (of good and evil) are considered as affects; that is, only insofar as they become affirmations of an increase in our power of acting. I take this to mean that the mere contemplation of a true idea by the mind is not yet an efficient cause at the level of affects,

but that it needs to be embodied in our activity. What it proposes as good has to be embodied in our activity; what it proposes as good has to be desired, i.e. actively striven for; and what it proposes as evil, must become the object of our contrary striving. In Spinoza's terms, the good is simply the name we give to what it is we desire, or what, in the desiring, gives us joy, or increases our power of acting.

The primary affects, then, characterize the principal modes of human action and interaction. They constitute, as it were, the internal psychological life of man, in virtue of his relation to other things requisite for his existence.

Man is, on this view, a creature of needs, and of vital activity in the service of these needs. His whole psychic life is the mirror of this activity, or its double in consciousness. However, it is not that consciousness *reflects* these needs, but rather that consciousness *is* identical with these needs themselves, under the attribute of thought.

What we have here, then, is a distinctive expression of the identity theory, not simply as a mind-body identity, but in this context, as an identity of thought and emotion, thought and joy, thought and sorrow, thought and desire. The radical consequence of this view is a rejection both of a mechanistic determination of psychic states by bodily states (an epiphenomenalist view too often mistaken for Spinoza's) and of a psychic determination of bodily states (the 'freewill' views against which Spinoza argues). For if one takes the identity seriously, as Spinoza constantly reminds us to do, every change in a psychic state is a change in a bodily state, necessarily; but not causally. A change in the psychic character, or intensity, or quality of an emotion does not *lead* to a change in a bodily state; it *is* one. Thus, the mistaken notion that Spinoza proposes a parallelism as against Cartesian interactionism simply has the model wrong. There is a conceptual parallelism, insofar as we think of bodies and of minds. But what we think, under these two attributes, is not parallel, but identical. The import of this identity for a theory of emotion, its complexities and difficulties aside, is that it refuses to assign the study of the emotions either to an exclusively physiological, causal-determinist model of explanation, or to an exclusively descriptivist, phenomenological, or teleological model of explanation.[8] Spinoza eschews mind-body interaction of the Cartesian sort not because he insists on an autonomy of (somehow) 'parallel' causal chains, but because he sees the emotions as variables in the dynamics of the life-activity of an integrated organism. The further import for a theory of

therapy, regarding the emotions, is that any therapeutic effect must involve a change in the life-activity of the individual, and not simply either a change in a psychic state, or in a bodily state. Spinoza's analysis of the emotions is not simply a way of understanding the emotions, but by virtue of this understanding, a way of dealing with them therapeutically. The normative construction is already contained in his very definitions of the primary affects, and of action and passion. This, therefore, is no mere anatomy of the emotions or passions, but rather a guide for the improvement of life, insofar as our life-activity itself is constituted by the dynamics of action and passion.

The whole construction of the psychology thus depends on the conception of a bodily organism – a 'complex body' or a 'composite of composites' – adequate in its complexity to feel, to suffer, to enjoy, and to think. That Spinoza intended just such a materialism of emotional and psychic life (or just such a psychic and affective capacity for organized matter) is, I think, clear. And nowhere clearer than in his own paraphrase of the Identity theorem in Part II of the *Ethics* (*E.,* II, vii) in the Scholium on actions and passions in Part III: "... the order of the actions and passions of our body is coincident in nature with the order of the actions and passions of the mind" (*E.,* III, ii, Schol.). Moreover, in what follows, Spinoza presents his striking argument for the capacities of the body, to account for the highest human activities, such as architecture and painting, in terms of bodily activities. Here, Spinoza develops his most powerful argument for a materialist mind-body theory, precisely in the context of what would be considered distinctively 'psychic' activities, and thus, presumably beyond explanation in terms of physiology. He does this in answering a skeptical doubt that the body, in itself, could have such elaborate capacities:

I scarcely believe . . . that, without a proof derived from experience, men will be induced to calmly weigh what has been said (about mind-body identity), so firmly are they persuaded that, solely at the bidding of the mind, the body moves or rests, and does a number of things which depend upon the will of the mind alone, and upon the power of thought. For what the body can do no one has hitherto determined, that is to say, experience has taught no one hitherto what the body, without being determined by the mind, can do and what it cannot do from the laws of nature alone, insofar as nature is considered merely as corporeal. For no one as yet has understood the structure of the body so accurately as to be able to explain all its functions . . . So that it follows that when men say this or that action of the body springs from the mind which has command over the body, they do not know what they say, and they do nothing but confess with pretentious words that they know nothing about the cause of the action, and see nothing

in it to wonder at . . . But my opponents will say, that from the laws of nature alone, insofar as it is considered to be corporeal merely, it cannot be that the causes of architecture, painting, and things of this sort, which are the results of human art alone, could be deduced, and that the human body, unless it were determined and guided by the mind, would not be able to build a temple. (*E.*, III, ii, S.)[9]

Now it would be puerile to interpret Spinoza as intending to argue here that the body is 'just as capable' of building a temple as is the mind, or that a body without a mind could do so. The polemic here is against dualism, and for identity. Nor is it for some statically conceived and merely analytical or formal identity of mind and body; rather, Spinoza is arguing for the identity of the *conatus*, the striving, self-preserving and self-enhancing life-activity of a minded body. It is the identity of a power of acting, and in this context, the emotions or the affects are crucial, as we have seen, since they are in effect nothing but the variations in this power. Thus, in the same Scholium, Spinoza says:

The decrees of the mind are nothing but the appetites themselves, which differ, therefore, according to the different temper of the body. For every man determines all things from his affect; those who are agitated by contrary affects do not know what they want, whilst those who are agitated by no affect are easily driven hither and thither. All this plainly shows that the *decrees of the mind, the appetite, and determination of the body are coincident in nature, or rather that they are one and the same thing, which when it is considered under the attribute of thought and manifested by that, is called a decree, and when it is considered under the attribute of extension and is deduced from the laws of motion and rest, is called a determination.* (*E.*, III, ii, S.) [My emphasis – M.W.]

If we are to understand Spinoza's theory of the passions at all, this identity is all-important. For it is at this (systematic) point that the continuity of a science of nature with a science of man is affirmed. Since, in effect, the science of nature is a search for the rational (causal) connections among bodies, by means of which we come to understand nature under the form of law, the extension of this program to the human sciences cannot introduce a different or alien methodological principle. The human mind and in particular the affections of the mind, its changes in its power of acting, now fall under a double condition: methodologically, the mind too can come to be understood rationally, i.e. in terms of its determinations; but this methodological extension has its warrant in the ontological claim: since the mind is 'nothing but' nature acting in a certain configuration, it is fundamentally no different from, and in fact, is identical with that nature which is also the subject of natural science; it is simply nature conceived under the attribute of

thought. The enterprise, therefore, becomes a "natural philosophy of the mind," or a science of psychology whose *identical* counterpart is a scientific physiology. There are thus not *two* sciences, but one; and this one is simply conceived under alternative attributes.

In this sense, Spinoza is not arguing for a reduction of psychology to physics, but rather for an extension or elaboration of the physics of bodies to take into account the activities of those complex and composite bodies which constitute persons or human beings; a 'physics' therefore, capable of explaining psychic function and human action. He uses the argument from complexity here, as a way of claiming for this complex organism those capacities for human action and art upon which his dualist opponents based their claims for free will and psychic agency. "I adduce also here the structure itself of the human body, which so greatly surpasses in workmanship all those things which are constructed by human art . . ." (*E.*, III, ii, Schol.). All this is therefore programmatic on Spinoza's part. His argument from ignorance (that we do not know the limits of corporeal activity) does not lead him, in the *Ethics*, to an actual examination of human physiology, which would begin to fulfill the program in actual research. Rather, he pursues his exploration at the level of psychological analysis, leaving it merely a programmatic claim that the "order of the actions and passions of our body is coincident in nature with the order of the actions and passions of the mind." I should rather say that this is more than simply a programmatic claim; for the metaphysical construction which Spinoza essays in the *Ethics* is made for the sake of supporting just this claim. And in this sense, I have argued that the metaphysics is for the sake of the scientific construction.

NOTES

[1] See my 'Metaphysics as Heuristic for Science,' *Boston Studies in the Philosophy of Science, III*, R. S. Cohen and M. W. Wartofsky (eds.), (D. Reidel, Dordrecht, 1965), pp. 123–170. Reprinted in this volume, pp. 40–89.

[2] This mereological principle is best expressed, perhaps, in Giordano Bruno's phrase, "wholly in the whole, and wholly in every part of the whole" (*De Immenso et Innumerabilibus, II*, xiii, cited by H. F. Hallett, *Aeternitas*, Clarendon Press, Oxford, 1930, pp. 155–156). We know that a related solution is sought by Leibniz, in the *Monadology*, and it shares much in common with Spinoza's; but it does not share Spinoza's insistence on the equal ontological status of matter with that of mind.

[3] See my earlier discussion of this point, in "Diderot and the Development of Materialist Monism," *Diderot Studies* II, (ed. by N. Torrey and O. Fellows), Syracuse University Press, Syracuse, 1952, pp. 279–327. (Reprinted in this volume, pp. 297–337. See especially,

p. 302.) There, I attribute mechanistic limits to Spinoza's view, based on the most explicitly mechanistic of the formulations in the *Ethics*, (*E.*, II, Axiom 2, Lemma 3, following Schol., Prop. xiii), and counterpose Leibniz, Maupertuis, and Diderot to Spinoza. I now think Spinoza himself strains and breaks the limits of this classical mechanism.

[4] Cf. Arne Naess, 'Freedom, Emotion and Self-Subsistence,' and Jon Wetlesen, 'Basic Concepts in Spinoza's Social Psychology,' both in *Inquiry*, 12, 1 (Spring 1969).

[5] However, see G. H. R. Parkinson's discussion of some difficulties in Spinoza's theory of falsehood and error: *Spinoza's Theory of Knowledge*, Clarendon Press, Oxford, 1964, esp. pp. 120–127.

[6] D. Bidney, *The Psychology and Ethics of Spinoza*, Yale U.P., New Haven, 1940, pp. 100–111; J. Martineau, *A Study of Spinoza*, Macmillan, London, 1895, p. 260 (cited in Bidney).

[7] Aristotle, *Eth. Nic.*, 1152b ff.; *Rhet.*, 1369b 33 ff.

[8] See, for a discussion on this point, L. S. Vygotskii, 'Spinoza's Theory of the Emotions in the Light of Contemporary Psychoneurology,' *Voprosy filosofii*, 1970, **6**, tr. E. E. Berg, in *Soviet Studies in Philosophy*, Spring 1972, pp. 362–382. This essay by Vygotskii, excerpted from the last chapter of a monograph bearing the same title, is part of the last of a seven-volume collection of his work, presently in preparation, which will also include 'Spinoza's Theory of the Passions.' Vygotskii contrasts Dilthey and Lange, among others, with respect to 'descriptive' (phenomenological) and 'explanatory' (causal) psychologies of the emotions.

[9] Bidney draws a very different conclusion from this passage, and one which I think is wrong. He interprets it to show that, for Spinoza, there are bodily affects which do not involve coincident affects of the mind; or, in effect, that Spinoza is inconsistent here, since elsewhere Spinoza argues that all bodily affections are also affections of the mind, on the identity thesis. Thus, if Spinoza admits affections which are purely corporeal, as he appears to, e.g. in his sleep-walking example (*E.*, III, ii, S.), or in his account of such affects as trembling, paleness, sobbing, and laughing (*E.*, III, lix, S.), the identity is violated. I think, on the contrary, that the point of the passage is that, for whatever mental affections there may be (especially those which we are prone to think of as *independently* mental), there are yet bodily affections coincident with them; that the body is therefore complex enough, in principle, to support such complex affections. Therefore, the thrust of the passage is not to assert that there are no mental affections independent of the body. Yet, as Bidney points out, Spinoza has no theory of unconscious affections with which to cope with those modifications or changes of the body which remain beyond conscious life. I think, rather, that what is missing in Spinoza's account is a notion of threshold, or emergent level, beneath which bodies are not minded, but *at* which they are. This is suggested, however, in Spinoza's notion of the mind as the idea of a composite or complex body. Spinoza is not a panpsychist, though the temptations to such an interpretation are great. Cf. D. Bidney, *op. cit.*, pp. 38–41.

NATURE, NUMBER AND INDIVIDUALS: MOTIVE
AND METHOD IN SPINOZA'S PHILOSOPHY

[1978]

I. INTRODUCTION

The problem I address in this essay is that of individuation in Spinoza's system. The apparent contradiction in the system is that Spinoza holds substance (Nature, God) to be a *unity*, to be *simple* (i.e. not compound, or composed of parts), *eternal* (i.e. uncreated and having no duration), and *infinite* (i.e. not determinate and not denumerable); and yet he also holds that 'in substance' there is infinite differentiation, there are determinate and finite modes, there is duration, and there are real individuals. Nor is it, for Spinoza, a case of there being both this *and* that: substance *and* modes, eternity *and* duration, one *and* many. Rather, he proposes that the infinity of modes is *identical* with substance, or is just the way substance is: a unity which is nevertheless infinitely differentiated; a simple which has no parts, but is nevertheless individuated; an eternal being which somehow expresses itself in duration; and an infinity which yet constitutes determinate, finite modifications of itself necessarily, in its activity.

Individuation has always been a crucial problem for interpreters and critics of Spinoza's philosophy. It was already, as we know, a central question in Spinoza's own discussion with his contemporaries. Thus, Tschirnhausen, among others, raised it sharply in his correspondence with Spinoza,[1] as did Oldenburg,[2] and Spinoza's well known response to Lewis Meyer[3] (The 'Letter on the Infinite') deals with the problem at some length. Similarly, in the series of letters to Huyghens,[4] Spinoza is again coping with the question of how the unity and indivisibility of God (or Nature) is compatible with the existence of individuals; and there too, as in his response to Jarig Jellis,[5] Spinoza is at pains to clarify his conception of number, in this context. The history of Spinoza interpretations is instructive as well. It divides rather neatly into those who try to hold Spinoza together, and those who want to tear him apart. And the issue over which this division most clearly takes place is that of individuation.

I am not proposing still another interpretation of Spinoza here, either as an attempt to discover the ultimate coherence of his system or to expose its contraditions. Rather, I want to accomplish two things: First, to show that the problem of individuation in Spinoza's system has its sources in the very way that Spinoza posed the question; second, to understand, historically, why he posed it in this way. That is, I want to reconstruct Spinoza's own *Problematik*: what question he posed for himself, and what motive he had for posing it; and further, to suggest the historical framework for the specific formulation of the problem, and the historical significance of Spinoza's profferred solution. In order to examine the problem of individuation in the specific and concrete form in which Spinoza treated it, I will also reconstruct the argument in Spinoza's own terms.

At the outset, then, I will consider some of the alternative frameworks for the resolution of the apparent contradiction in Spinoza's system. Then, I shall propose a thesis as to what Spinoza's motive may have been, in formulating the problem in a certain way. Here, I will discuss what I take to be the integral relation of Spinoza's method to his motive, and attempt to reconstruct the philosophical rationale for the method. My proposal here is that, in effect, ontology recapitulates methodology, in Spinoza's construction, and that one cannot understand the ontology itself unless one understands both method and motive. Further, I will argue that the sources of the apparent contradiction in the system lie in Spinoza's motive and method. I shall then examine some of the details of Spinoza's argument on individuation, specifically, his concepts of *individual*, and of *number*, in relation to his account of the three kinds of knowledge. And, finally, I shall suggest what the historical framework is for Spinoza's reflection, and what historical problem it is to which Spinoza's philosophy offers an attempted solution.

II. ALTERNATIVE RESOLUTIONS OF THE APPARENT CONTRADICTION IN SPINOZA'S SYSTEM, AND A PROPOSAL FOR A NON-RESOLUTION

Interpreters and critics of Spinoza have offered either alternative proposals for resolving the apparent contradiction in Spinoza's system, or alternative ways of revealing its incoherence. Spinoza's account of individuation leads to the following apparent contradiction, to restate it briefly here: Spinoza holds that substance (God or Nature) is a unity, is

simple, eternal and infinite. Yet he also holds that 'in' substance, there is infinite differentiation, there are determinate and finite modes, i.e. there are individuals. The simplest solution to this problem has been to claim that only substance is 'real' (in its unity, simplicity, etc.) and that attributes have no ontological status, but are simply 'ways' of conceiving substance; or to claim that the attributes have the ontological status of such 'ways' only, and are therefore limited to the being of conception, or of thought (with all the attendant difficulties of that view: For example, the attribute of *extension* 'exists', therefore, only as a way of conceiving substance). On such a reading, the modes of substance and, in particular, the finite modes, are not 'real', and are merely appearances, or, in the worst case, are affirmed to exist only in consequence of our inadequate ideas, and therefore have the status of errors in judgment, or fictions of the imagination.

An alternative solution to the problem is to interpret the modes, ontologically, as 'deriving' their reality from substance, in some transitive and productive causal relation. This issues in one or another form of the conception of grades or degrees of reality. In fact this is consistent with Spinoza's explicit view that there are indeed grades of reality, or of perfection; but runs into the difficulty of understanding how each of these grades yet 'expresses' the infinity and perfection of substance, not partially, but wholly.

Still a third interpretation, and perhaps the most difficult to maintain, is that for Spinoza only individuals are real; and that to say that substance is identical with its modes, and is an immanent rather than a transcendent cause, as Spinoza says, is to say in effect that substance is *nothing but* the totality of these individuals, understood in their necessary connections with each other. Substance, therefore, is simply the system of individuals (which system is not produced by anything outside it, but simply *is*), and therefore, that the system as a whole is *causa sui*, though it has no being apart from its individuals.

All of these attempted resolutions of the apparent contradiction in Spinoza are to my mind inadequate, though each has a warrant in *some* of what Spinoza says. The more promising approach seems to me to be that which takes Spinoza at his word, and refuses to accommodate him either to Platonism, neo-Platonism or nominalism. What does it mean to 'take Spinoza at his word'? It means, I belive, to understand him as saying that substance is *identical* with the infinity of modes, that it is a unity which is nevertheless infinitely differentiated; a simple without

parts which, yet, is individuated; an eternal being which expresses itself in duration; and an infinity which constitutes determinate, finite modifications of itself in its activity. In short, to take Spinoza at his word is to acknowledge the contradiction, but to refuse to resolve it either by a Platonist interpretation, i.e. by *ontological elimination* of the modes, construing them as mere appearances; or by a neo-Platonist interpretation, i.e. by *ontological attenuation* of the modes, seeing them as deficiently real, or as partial expressions of substance; or by a nominalist interpretation, i.e. by *ontological abstraction*, in which substance is a bloodless universal riding herd on bloody particulars.

Why is it more promising to take the contradiction head on, and refuse to resolve it? Because, it seems to me, it leads to an attempt to understand why Spinoza posed the question of substance and modes as he did, and why this very way of posing the question gives rise to a contradiction. The contradiction, on this view, is fruitful: it is not judged simply as Spinoza's failure, in formal terms, to achieve a philosophically consistent construction. Rather, the contradiction is seen as a clue to understanding Spinoza's philosophy. The roots of the contradiction may then be shown to lie in the very character of this philosophy, as a radically transitional one, bridging two incompatible systems of thought and, indeed, two incompatible worlds.

In order to understand the character of this philosophy, then, I want to begin by examining Spinoza's motive for his construction, and its relation to his method.

III. MOTIVE AND METHOD

When I speak here of Spinoza's motive in the construction of the *Ethics*, I am not adducing his psychological motive, his unconscious or subconscious motive, nor yet the ways in which his social and political context may have shaped his motive. Rather, I am attempting to reconstruct Spinoza's own motive as he himself may have understood it; that is, his motive as the conscious, or self-conscious aim of his work. I do not mean, by this, to deny the importance of psychological or socio-political contexts. I am simply arguing that, in order to address such questions in any serious way, we have *first* to explicate and to understand how Spinoza himself formulated his problem, and how he proposed to resolve it.

Spinoza's motive was intellectual. That is to say, his motive was the

satisfaction of the intellect. What the intellect is, and what would satisfy it is not for us to say, in this context. Rather we must let Spinoza tell us what he takes the intellect to be, and what he takes to be its satisfaction. In short, we have to begin by accepting Spinoza's own understanding of what he took to be the highest activity for human beings, and his own claim that, in the *Ethics*, he was himself engaging in that activity.

Here we are on fairly solid ground. For Spinoza tells us in great detail what he takes the intellect, or the mind, to be. The intellect, or the mind, is the activity of coming to know, it is the activity of thought, whose end is knowledge. What satisfies it is knowledge. Its being, then, is knowing. The more it comes to know, the more being it has. Its perfection, so to speak, is perfect knowledge. But knowledge of what? Perfect knowledge cannot be knowledge of this or of that determinate object of knowledge. It cannot be perfect only with respect to a limited or finite domain, for then, there would be *more* perfect knowledge, which goes beyond these limits. Therefore, finite knowledge cannot be perfect. On Spinoza's view, then, perfect knowledge must be infinite knowledge of an infinite object. Infinite knowledge of a finite object already contains the conception of the finitude of its object, i.e. it knows itself to be limited, if it knows its domain to be limited, even if everything in that domain comes to be perfectly known. The object of perfect knowledge must, therefore, be an infinite object. Only such an object can satisfy the very nature of the mind as an activity of coming to know. Further, there cannot be more than one such infinite object, for if there were two or more such objects, each would be bounded by, or delimited by the other. Nor can such an infinite object have its own source, or ground in something beyond itself, for then again, its being would depend on something else, and it would, in this sense, be finite (created or generated). Our knowledge of it would depend on the knowledge of what lies beyond it, and it would therefore not be, in itself, the appropriate object for the satisfaction of our knowing-activity.

In some such way, one may reconstruct how Spinoza may have postulated the very condition for the activity of the intellect as the existence of that infinite being or substance, *Deus sive Natura*, which alone could provide for the satisfaction of the mind.

It should be clear by now, to readers of Spinoza, what sort of game I am playing here. In reconstructing Spinoza's motive as intellectual, and following his own account of what the intellect is, I have *derived* an account of what Spinoza's substance *must* be, if it is to be the

appropriate object of thought, or what alone could satisfy the intellect, or realize its nature. It may be objected that my account of Spinoza's motive is topsy-turvy; that in starting from the intellect and its characterization, I am ignoring Spinoza's own construction, which begins instead with substance. In effect, it may be objected that I am reversing the order of Parts I and II of the *Ethics*; or, indeed, that I am beginning with Part V ('of the Power of the Intellect, or of Human Freedom'). In fact, I am proposing just such a thoroughgoing reversal of the order of the parts of the *Ethics*, in order to understand what Spinoza's motive was. What I propose, then (as I have argued elsewhere),[6] is that the metaphysical construction in Part I ('Of God'), and the systematic construction of a world system in Part II ('Of the Nature and Origin of the Mind') are for the sake of the psychological and ethical theory in Parts III–V. In effect, then, the motive for Spinoza's philosophical construction, in general, lies in Part V. The argument of the work, construed in this way, concerns the conditions under which human freedom is possible, where this freedom is identified with the activity of the intellect, as Spinoza defines it.

When I say that Spinoza's motive is intellectual, then, I am claiming that Spinoza took the satisfaction of the intellect (as he understood it, of course) to be the aim of his work. But in order to take this motive in its full seriousness, we need to understand that the activity of the intellect is, for Spinoza, the highest activity of which human beings are capable, by their nature; and that human freedom, for him, consists in the fullest realization of this essential activity. Nor are we to regard this activity as one of quiet and detached contemplation, but rather as involving itself in the very subject which is alone adequate to it: God or Nature. The fulfillment of the intellect's own nature thus involves the scientific understanding of nature, the rational organization of the state, and the satisfaction of human desires in the way most optimal for the preservation of life and the enhancement of our happiness. For Spinoza, the intellect is no neutral observer, but rather activity.

It would seem that, in putting things in this odd, or reversed way, I am in fact violating Spinoza's own account. Where he begins with God or Nature, and derives the origin and nature of the mind and the possibility of human freedom, I propose instead that we construe him as beginning with the intellect and human freedom, and deriving the nature of substance therefrom. The difficulty may lie in the notion of 'derivation' here. Spinoza's derivation, in the *Ethics*, is by way of deduction of

propositions from axioms and definitions. The 'derivation' I am proposing is, in effect, a transcendental deduction: *given* the nature of the intellect and human freedom, what would have to be presupposed as necessarily existing such that the intellect thus defined could realize its nature? To put the question in a different way: How is knowledge possible? Or how is the attainment of knowledge (i.e. the very activity which defines the intellect) possible?

Why is this a clue to the understanding of Spinoza's formulation of the question of individuation? And how does it help us to see the sources of the contradiction in Spinoza's system in the very way in which the question is formulated? Here, too, I want to propose that we reverse the usual formulation, in order to realize how Spinoza's motive itself affects the issue. The usual way of posing the problem of individuation, or the existence of individuals, or finite modes, is to ask how infinite substance gives rise to finite individuals, or how the One produces or generates the Many. Suppose, however, we formulate the question thus: How can a finite individual attain the knowledge of substance? Or, to put it differently, how can the human mind, which is on Spinoza's account an individual thing, have as its *conatus* the knowledge of infinity? Spinoza himself appears to block the path of such reasoning, in his insistence that the infinite cannot be derived from the finite, and certainly not by summing the totality of finite modes. Yet, according to Spinoza, the mind *can*, by its very nature, know God, know that he necessarily exists, etc., etc. The motive of infinite knowledge requires, for its realization, that an individual can come to know infinite substance. But this is possible only if the individual mind is capable of infinite knowledge, and is therefore itself infinite in its capacity. The source of the difficulty, put in this way, is not: How can the infinite know the finite? nor yet, How can the infinite know the infinite? but rather, How can a finite individual, defined as a mode of substance, come to have an infinite capacity? We may also rephrase the question, in terms of Spinoza's system, as one which asks: How can a determined being be free? For finite modes are fully determinate, and yet one of them (the human being, as a thinking being) is capable of freedom.

The force of posing the question in this way is that it begins with existing individuals and with their capacity for freedom. Moreover, their capacity for freedom lies in their very nature as thinking beings, for it is in the practice of thought, by the right method, that freedom may be realized (and this is the burden of Parts III–V of the *Ethics*). Yet,

Spinoza insists that substance is ontologically prior to its modifications (though not as transcendent, but rather as immanent cause); and, further, that it is *identical* with these modifications, though it produces them. Thus, though we *begin* with individuals on the way up (from the existence of the human mind to the presupposition that it requires infinite substance as its object), we *end* with individuals on the way down (from the necessary existence of substance to its self-individuation or modification).

The source of contradiction, in this (reversed) form of posing the problem of individuation, is thus related to Spinoza's motive itself. The motive is to show that finite individuals (human beings, or Spinoza himself) can know God or Nature; that is, that the world is fully intelligible to the human mind, without residue; and that finitude is therefore *somehow* compatible with freedom.

Spinoza's method is the correlate of his motive. I can only briefly set forth here what the correlation is. If Spinoza's motive is the satisfaction of the intellect, i.e. if his motive is to show that an individual mind can attain to infinite knowledge, then the method must be one by means of which such knowledge can be attained. But the issue is somewhat trickier here. For I have argued that the very metaphysics which Spinoza constructs is such that God or Nature turns out to be necessarily fully intelligible to the human mind. The method, therefore, simply defines intelligibility. That is, the method is an account of the nature of the intellect, as Spinoza understands it. And, therefore, the intellect, in operating according to its own nature, *must* attain to its end of infinite knowledge.

Let me put this still another way. One would ordinarily suppose that the method is determined by the nature of the object of inquiry. Thus, the question of method would be: Given the nature of substance (God or Nature), how could it come to be known? However, I am proposing that Spinoza's approach, in fact, is different. Namely, it proceeds from the nature of the intellect, and not from the nature of substance; and this is because, as I have argued, Spinoza's construction of the nature of substance itself proceeds from a consideration of what would be required as the adequate object of knowledge for the intellect, defined in a certain way (i.e. as perfect, infinite, etc.). In short, Spinoza's methodology is not developed on the basis of a prior ontology, but rather, the ontology is constructed on the basis of a methodology; and the methodology in turn, is constructed as a definition of the way in which

the intellect operates, by its nature. That is, method defines the natural activity of the mind, and since the mind *is* its mode of acting, method defines the very being of intellect, and thus the very intelligibility of God or Nature. Method is linked to motive, then, precisely in that the method simply mirrors the activity which, by definition, would satisfy the intellect.

Spinoza, of course, does not put it this way. But the way in which he puts it is revealing. Let me summarize it briefly thus: the right method is that which is in accordance with the nature of an infinite intellect, or (what comes to the same thing), in accordance with the intelligibility of God or Nature, or (what comes to the same thing again), in accordance with the very essence of God or Nature, namely, that it knows itself infinitely and absolutely. It follows, then, that the method which Spinoza proposes for our understanding of God or Nature is God's (or Nature's) *own* method of knowing itself. The epistemological force of this claim cannot be over-estimated. For how could God's own way of knowing himself be mistaken? And how could we help but arrive at a knowledge of God, *necessarily*, if we follow this method itself?

What Spinoza offers us here, therefore, is not itself the content of our certain knowledge, but the conditions for the possibility of certainty, namely, God as a model of the human intellect; or, to put it differently, an argument to the effect that God knows himself through our knowledge of him; that the human mind is, in its essence, God's mind, in its infinite modifications under the attribute of thought; in short, that the intellect is divine. Method is therefore God's own way of knowing himself; or, less theologically, method is Nature's way of knowing itself. And since 'the order and connection of thoughts is the same as the order and connection of things' (II P7),[7] this infinite knowledge is at once a knowledge of body and of mind, or of Nature conceived of under the attributes of extension and thought.

IV. INDIVIDUALS

Now we come to the crux of this essay. The problem I posed at the outset was that of understanding how the contradiction in Spinoza's system was rooted in his formulation of the question of individuation. I proposed that, in order to understand Spinoza's formulation, we needed to grasp his motive and its relation to his method. Now, if the motive is, as I have suggested, the satisfaction of the intellect, and further, if the

satisfaction of the intellect is that activity which realizes its nature (i.e. coming to know), then the object of its activity must be adequate to the satisfaction of this nature, i.e. it must be infinitely knowable. On this basis, I suggested, Spinoza constructs this object. God or Nature, in accordance with this requirement. What follows from God's nature, as the object of infinite intellect, must be such that it provides the conditions for the satisfaction of this intellect.

Now Spinoza has a problem: on this model, the infinite intellect, or knowing mind, exists; and its object exists as an infinite object. But, what is infinite is unbounded, i.e. it cannot be delimited. And, therefore, two infinite things cannot exist, for one would delimit the other. The intellect therefore must be identical with its object; or, to put it differently, the intellect must have itself as its own object, or the infinite object must be self-knowing. On this model, then, there exists at most only one thing (substance) which knows itself infinitely. What is it, then, that is known? The activity of knowing itself, as an undifferentiated entity.

But, for Spinoza, such abstract self-contemplation cannot be coherent with another fact: that it is the human individual who thinks, and that human individuals are finite in existence. Moreover, thinking involves thinking *this* and *that*. Thinking is not thinking one thought, but different thoughts. Further, the thing thought is itself an individual, as the object of an individual thought. Yet, it is in this activity of thinking that the nature of thinking is satisfied; and the requirement for the infinity of thought can only be satisfied if the totality of thoughts (and of things thought) constitute such an infinity. But, as we know, Spinoza argues that an aggregate cannot constitute an infinity, or, to put it differently, the infinite is not constituted of parts of itself, nor is it denumerable.

Two alternatives present themselves: either there is only one individual, and there are no finite existences or individual things (individual thoughts and things-thought); or there is no infinite substance, but only the aggregate of finite individuals. Spinoza doesn't accept this *either-or*, however. Therefore, the crucial question is: what must individuals be like so that they are the adequate objects of an infinite intellect? Spinoza's solution, as we know, is that individuals are such that each of them 'expresses' the infinite. But what could this mean? Clearly, each individual can express the infinite only if the individual is infinite. But we cannot have a plurality of infinites, on Spinoza's view.

Spinoza's way out is to argue that individuals must be such that they can exist, or be known only *through* the infinite; the infinite must be known *in* them, so to speak, and therefore, they cannot be known in themselves. What is it, however, that is known *in* them, as individuals? Simply, *that* they cannot be known in themselves; i.e. to know *that* they cannot be known in themselves is to acknowledge, at one and the same time, that they are finite and that something else exists in itself with respect to which their finitude is determined. In short, the concrete way in which the infinite becomes the adequate object for an infinitely-knowing mind is in the recognition of the finitude of individuals. The necessary existence of infinite substance is thus the presupposition entailed in the very conception and existence of finite individuals, as the ground of their existence; and the knowledge of the finitude of these individuals is the way in which this presupposition is concretely realized in thought.

The knowledge of the finitude of individuals, however, is nothing but the knowledge of their determinateness, their individuality itself. It is knowing what they are, essentially. But if their finitude or determinateness is what constitutes them as being what they are essentially, the knowledge of each individual entails the knowledge of its particular or singular determinations, namely, its individual essence. In knowing its individual essence, we come to know it not simply in the context of its connections with all other individuals, each equally determinate, but we come to know that its finitude, determination, singularity – namely, its individuality itself, its being *that* individual and not another – is the expression of the infinity with respect to which it exists as that determinate individual and no other.

That this argument in Spinoza is notoriously difficult, and that it is finally unconvincing is beside the point here. It is an attempt to mediate the infinite and the finite, in the context of a project whose proclaimed end is that the knowledge of individual things *is* the knowledge of God or Nature; and that such knowledge is possible.

Now we need to step outside the argument for a moment, to see it in relation to Spinoza's motive and method. The question of whether there are individuals is not at issue for Spinoza. Nor is the question whether knowledge is possible, or whether individuals are free. The question, therefore, is not whether these things are *true*, but whether they can be *proved*. The existence of individuals, of knowledge, of freedom, are not taken as hypotheses by Spinoza, but rather as conclusions already known. What he needs to do, in his construction, is to provide the

presuppositions from which these conclusions follow deductively. The test of Spinoza's system is not whether his conclusions are true, but whether the system is adequate to arrive at these conclusions deductively, i.e. whether Spinoza's concepts of God or Nature, or of individuals are such that what is known to be true follows from them. What Spinoza knows to be true, then, and what he takes to be the aim of his proof is to be looked for at the end of the construction, and the construction itself understood in the light of this aim. If we cast the *Ethics* in this light, as I propose, then what is at issue for Spinoza is the formulation of a theory of individuation adequate to the task of proving rationally, i.e. by deductive inference, what is already known on other grounds, i.e. intuitively.

Here, I will briefly propose such a reconstruction of Spinoza's theory of individuation. Again, we should begin with Part V of the *Ethics*, rather than Part I, in this reconstruction. In Part V, Propositions 24–33, Spinoza considers the 'third kind of knowledge', namely that which he defines as 'intuitive science': 'This kind of knowing advances from an adequate idea of the formal essence of certain attributes of God to the adequate knowledge of the essence of things' (II P40 Note 2). This, it seems to me, is the central claim Spinoza makes for the possibility of knowledge, and he couldn't have made it more strongly. For what is involved, in the notion of the third kind of knowledge, is the knowledge of the essence of individual things, *and* the notion that such knowledge follows from knowledge (or an 'adequate idea') of 'the formal essence of certain attributes of God'. Here, the claim is that knowledge of individuals proceeds from knowledge of infinite substance.

Spinoza is careful in his insistence that such knowledge of individuals does *not* proceed from the second kind of knowledge, i.e. from common notions; for, while common notions yield adequate ideas, they yield adequate ideas of what we may call nomic universals, and which Spinoza characterizes as "those things which are common to everything, and which are equally in the part and in the whole . . ."(II P38). From such common notions, however, nothing can be deduced concerning the nature of individuals, except what is abstractly or generally true about them. Thus, Spinoza writes, "That which is common to everything and which is equally in the part and in the whole, forms the essence of no individual thing" (II P37). That is to say, the nature of the individual, for Spinoza, is not constituted by common properties, but distinctively, by its unique properties. The essences of individuals are individual essences. And these can be known only by the third kind of knowledge.

Further, Spinoza obviously excludes the first kind of knowledge by which we know individual things only as "they are represented by the senses to us in a mutilated and confused manner and without order to the intellect", or, "from signs . . . through which ideas we imagine things" (II P40 Note 2). This kind of knowledge, which Spinoza calls opinions, or imagination, includes the 'indistinct' images formed by the mind which are denoted by so-called 'transcendental terms' (e.g. *being, thing, something*) and what Spinoza calls "universals" (e.g. *man, horse, dog*). Moreover, those things which are aids to the imagination, which simply help us to order the flux of perceptual experience, but which are not essential to the nature of the things known in this confused way, are *entia rationis*, 'things of reason', and have no necessity to them. These are *time, number,* and *measure.*

The third kind of knowledge, by which alone individuals can be known, is the highest kind of knowledge, for Spinoza. In terms of the argument I have offered earlier, it is this kind of knowledge alone which can satisfy the intellect or realize its nature. This is the burden of Spinoza's argument in Part V, Propositions 24–33 (especially; elsewhere, too, of course). Proposition 24 states: 'The more we understand individual objects, the more we understand God', and Proposition 25 adds that it is by the third kind of knowledge of the essence of things that the mind exercises its highest virtue, i.e. its power of acting or its nature. Proposition 27 and its Demonstration state in the clearest and most striking way Spinoza's conception of the activity and aim of the intellect, and may be taken, I believe, as the paradigmatic statement of Spinoza's project. I quote it in full:

From this third kind of knowledge arises the highest possible peace of mind. Demonstration. The highest virtue of the mind is to know God (IV, 28) or to understand things by the third kind of knowledge (V, 25). This virtue is greater the more the mind knows things by this kind of knowledge (V, 24) and therefore he who knows things by this kind of knowledge passes to the highest human perfection, and consequently (Def. 2 of the emotions) is affected with the highest joy which is accompanied with the idea of himself and his own virtue (II, 43); and therefore (Def. 25 of the emotions), from this kind of knowledge arises the highest possible peace of mind. – Q.E.D.

What is the force of Spinoza's assertion here? It is not that the human mind can realize its perfection *either* by knowing God or Nature, *or* by knowing individual things, but that knowing individual things *is* the knowledge of God or Nature. Therefore, the human mind cannot achieve its perfection, i.e. it cannot act in accordance with its nature unless there

are individuals knowable in this way. To be an individual knowable in this way, however, is to be a unique individual, i.e. not an abstract particular, but a concrete individual. To be concrete in this way is to stand in relation to all other individuals as really distinct from them (i.e. not simply in thought, but in existence), and therefore to be understood in connection with the system of individuals, not simply as a part of that system, but as expressing, in its very individuality, i.e. its difference from all the others, or the qualitative infinity of the system itself.

The theory of individuation that Spinoza proposes, then, has to satisfy the requirements for knowledge of a certain kind, i.e. knowledge of qualitatively distinct individuals – i.e. individual essences – and it also has to satisfy the requirement of systematicity, i.e. that these individuals are necessarily connected. This latter requirement derives from the nature of the intellect as well, since an adequate idea of an individual requires the knowledge of its determinations, i.e. either its self-determination, if it can be conceived in itself, or has the ground of its existence in itself; or its determination by what lies beyond it, if it has to be conceived of or has the ground of its existence through another. But, according to Spinoza, simply to conceive one individual in its determination by another, in the case of finite modes of substance, is to conceive it in its contingency, and not yet in its necessity. The conception of individuals in their necessity must therefore be grounded in the conception of the whole system of individuals as necessary, namely from the point of view of the infinite whose necessity lies in its having nothing beyond it to determine it, and which therefore cannot be other than it is, i.e. other than as it determines itself. The non-contingency of the infinite, i.e. of substance, is precisely in its depending on nothing beyond itself for its existence.

In Spinoza's view, then, the human mind approaches such an understanding of the necessary order and connection among individuals in knowing concrete individuals, and in knowing more of them more and more concretely, both with respect to their uniqueness or difference and with respect to their relations to other individuals. These two respects are systematically complementary, however, for to know an individual essence is to know its individuality not through itself alone, but by virtue of the infinite differences between it and all other individuals. If, as Spinoza says, every determination is a negation, then in fact the very uniqueness of an individual consists as much in all the ways it is *not* like any other individual, as in the ways in which it is what it is. For every

property or quality an individual has, which makes it that individual and no other, there is a property or quality which some other individual must have which distinguishes it from the first. That is to say, there are no free-floating or uninstantiated properties for Spinoza. To be a property (or a determination, to use Spinoza's term), is to be the way some individual is. And since, for Spinoza, the intelligibility or knowability of the object of knowledge is inexhaustible, if it is to be adequate to the infinite intellect, then there are infinitely many qualitatively different individuals. We may now read Proposition 16, Part I, in this light: "From the necessity of the divine nature infinite numbers of things in infinite ways (that is to say, all things which can be conceived by the infinite intellect) must follow." But we may interpret it as saying, in effect, that if the human mind is to be able to realize its nature, then there must be such an infinity of qualitatively distinct individuals for it to know.

On this construal of Spinoza's theory of individuation, one would then read whatever Spinoza says about individuals as part of a systematic construction whose end it is to prove that the human mind is not finite in its capacity for knowledge, and that only confused ideas lead to the notion of its finitude.

What are these confused ideas? Spinoza's theory of knowledge gives us a systematic account of them, and of their origin, and I do no propose to review it here. But one of these confused ideas has bearing on the question of individuals, and that one may be examined briefly here, to establish the spirit of Spinoza's criticism. This confused idea or 'thing of the imagination' is *number*.

V. NUMBER

The difficulties in Spinoza's system are nowhere greater than in his discussion of number, for here the issue of the relation of individuals to substance appears in one of its sharpest forms. It would initially appear, from Spinoza's account, that if there are real or concrete individuals, such that they are individual essences, then they must be denumerable, one by one. The identity of individuals, it is true, is a matter of their unique determination in the system of individuals, and thus involves the relation of each individual to all the others. Nevertheless, this relational character of the individuals does not subvert the distinctness of each individual, for it is precisely in terms of this distinctness that Spinoza

argues that individuals are known by the third kind of knowledge; and also, on these grounds, that common notions, or the second kind of knowledge, cannot individuate.

It would then appear that, since the mind attains its perfection by increasing its knowledge of individuals, and since there are infinitely many individuals, there must also be some corresponding arithmetic notion of number which derives from the individuality of individuals, and from the fact that there is a plurality of them. Thus the mind's conception of number should, one would expect, proceed from the notion of *an* individual, as a unit, to the notion of two individuals, as distinct units, and thus to the notion of several individuals, as an aggregate or sum of such units, to the notion of indefinitely many, i.e. to the notion of an unending or continuing sum of such units, with no upper bound. Number, conceived of here as what can be counted, pertains therefore to entities which are distinct individuals or things, and of which there can be more or less. So it would seem.

But Spinoza ascribes this whole arithmetic conception of number to a confused idea, deriving from the weakness of the imagination, which is bound to the limits of what can be perceived, or remembered. When Spinoza talks about 'infinitely many' individuals, he is not talking about some 'greatest number'. Rather, he argues, the concept *number* is totally and categorically inapplicable to an infinity, and can be applied to it only by a fundamental misconception of the nature of the infinite. 'Infinitely many' is therefore not to be taken as 'indefinitely many', but as denoting what is non-denumerable. 'Non-denumerable' here is not meant therefore as beyond enumeration on the grounds that we do not have enough numerals, but as fundamentally and everywhere non-denumerable. Thus, there are not *some* individuals in the infinity of them that are denumerable, while others are not; but, rather, individuals are *in themselves* not denumerable, if we understand rightly what they are, i.e. if we know them by the intellect, and by the third kind of knowledge, rather than by the first kind, i.e. by imagination, as vague or confused ideas.

The *local classicus* of Spinoza's discussion of number is in his reply to L. Meyer (the so-called 'letter on the infinite').[8] Here, the discussion distinguishes what we may say about modes or individuals taken in themselves, apart from substance, and what we may say about them when we conceive them in their relation to substance, i.e. as its modifications. The concept of number arises only in the former case,

when we deal with individuals abstractly, i.e. abstracted from their real being 'in' substance. Thus, Spinoza writes:

> . . . from the fact that we can limit duration and quantity at our pleasure, when we conceive the latter abstractedly as apart from substance, and separate the former from the manner whereby it flows from things eternal, there arise *time* and *measure; time* for the purpose of limiting duration, *measure* for the purpose of limiting quantity, so that we may, as far as is possible, the more readily imagine them. Further, inasmuch as we separate the modifications of substance from substance itself, and reduce them to classes, so that we may, as far as is possible, more readily imagine them, there arises *number*, whereby we limit them. Whence it is clearly to be seen that measure, time, and number, are merely modes of thinking, or rather, of imagining.[9]

Here, as elsewhere, Spinoza characterizes *number*, together with *time* and *measure*, as *entia rationis*, and as aids to the imagination, but as fundamentally false, if they are taken to be characteristics of substance, *or* of the modes adequately understood, as they should be, in relation to substance.

The error arises, says Spinoza, when we impose upon the infinite the limits of the imagination, or the conveniences of 'abstraction' (i.e. taking things apart from their relations to the infinite totality). It would then appear that individuation itself – the distinctness and unique identity of individuals – is just such a confused idea, a division of the infinite into parts, which Spinoza argues cannot be done without destroying the infinity itself. But what else can the individuals be? Or what must individuals be if they are not parts of a whole, and not denumerable, and are yet individual essences?

In some places, Spinoza talks about the distinction between concrete individuals (i.e. those comprehended 'under the aspect of eternity') and abstract individuals (i.e. those taken apart from substance) as simply two ways of conceiving of them. Thus, in the same letter to Meyer, he writes that:

> [C]ertain things are in their nature infinite, and can by no means be conceived as finite; whereas there are other things, infinite in virtue of the cause from which they are derived, which can, when conceived abstractedly, be divided into parts, and *regarded as finite.*[10] (Emphasis mine)

Are finite modes then no more than fictions of the imagination, and not real? Are they simply finite in being 'regarded as finite'? So it would seem. Yet, our knowledge of God or Nature proceeds through the knowledge of individual things; and if these individuals were less than real, by virtue of their individuation as finite, determinate entities, or if

these individuals were individuated only as modes of imagining, then our knowledge of the infinite would proceed from knowledge of the finite, or on the basis of inadequate ideas, i.e. of imagination or knowledge of the first kind. But this is not permitted, by Spinoza's own argument. The only resolution possible, if Spinoza is to maintain that knowledge of the third kind is knowledge of individuals, and also that knowledge of this kind is knowledge of the infinite, is to propose that individuals themselves are infinite. But since there can be only one infinite, and there are infinitely many individuals, each one cannot be infinite, for then there would be a plurality of infinites, or, in effect, infinitely many infinites. Individuals, if they are infinite in any sense, cannot be denumerable in respect to their infinity, therefore. In what respect can they be infinite, then?

Spinoza's well known argument here is that an individual is infinite, or is conceived by the mind as infinite, insofar as it is conceived in relation to God as the source of its being.

But since any individual or finite mode is defined as having the ground of its existence beyond itself, and, in substance, no existing thing can be conceived apart from this relation to substance; and insofar as it is so conceived, the idea of its existence also necessarily involves the idea of substance, or of the eternal and the infinite. In effect, every proper conception of an individual thing *reminds* us of its ground in infinite substance. Thus, though individual things are not themselves infinite, the right way of thinking about them implies the infinity of substance as their cause. Similarly, the individual human mind is eternal and infinite insofar as it conceives itself and its body under the form of eternity (V P30, P31). Further, Spinoza writes:

[T]hings are conceived by us as actual in two ways – either so far as we conceive them to exist with relation to a fixed time and place, or insofar as we conceive them to be contained in God and to follow from the necessity of the divine nature. But those things which are conceived in this second way as true or real we conceive under the form of eternity, and their ideas involve the eternal and infinite essence of God . . . (V P39 Note)

The simplest form of the argument, then, is that substance is in all things, as their immanent cause. But it is not divisible, i.e. parts of it are not distributed in all things, since it has no parts and is not a composite. And, therefore, things considered in themselves *may* be taken as parts of a whole, and as distinct entities, subject to enumeration; but taken in their true sense as modifications of substance, are not denumerable as units, additively. Yet they retain their individuality, for otherwise they would

all become indistinguishable from one another, and vanish into the One. The imagination distinguishes things numerically, in order to limit the infinity, and to help order things for everyday purposes. The intellect cannot limit the infinity without destroying or reducing its infinite object, and thereby limiting itself, as an infinite activity of knowing. Therefore, the intellect sees past number to the nature of the true infinite, and realizes this infinity in and through its knowledge of individuals, insofar as these are conceived under the form of eternity, i.e. in their derivation from God or Nature.

   This is, at the very least, a clever arugment, and at its best, a profound idea about the relation of individuals to substance. But it smells of compromise or contradiction, since it either posits individuals which are not really individuals, i.e. only appear to be individuated to the imagination; or it posits finite individuals as somehow infinite in their essence, yet singular in their existence. The notion of real individuals which are nevertheless not denumerable exhibits the enormous tension in Spinoza's system between the concept of an infinite nature and the concept of individuality.

VI. SOME REFLECTIONS ON THE HISTORICAL CONTEXT OF THE CONTRADICTION IN SPINOZA'S SYSTEM

From this brief reconstruction of Spinoza's theory of individuation, and its problems, we may turn finally to a consideration of the reasons why Spinoza may have formulated the problem in this way, and why he resolved it, or failed to resolve it as he did. There are, of course, dialectical problems in any thoery of individuation, and one may approach them (as most Spinoza studies do) in a purely conceptual way. Similarly, since Spinoza inherited and made use of a rich and complex background of philosophical discussion on these issues, if one were doing an internalist history of ideas (say, like Wolfson's study of Spinoza) one could no doubt show the deep historical continuity of the philosophical *Problematik*. And one could conclude that Spinoza simply worked out a major alternative conceptual formulation of the problem.

   But there is yet another important context for the understanding of Spinoza's work, and that is his own time, his historical period. Hegel's dictum that philosophy is its own time comprehended in thoughts is perhaps too restrictive, too retrospective a framework for the full understanding and evaluation of a philosophical system, and may keep

us from recognizing the contemporary significance, or from giving contemporary interpretations of Spinoza's thought. Yet, setting a philosophical problem into its historical framework may help us to understand its larger significance. If we understand Spinoza's own motive, the satisfaction of the intellect, in its own time and place, then it is not simply an abstract or idiosyncratic motive, and certainly not a narrowly philosophical motive. Moreover, if we read the *Ethics* against the background of the *Theologico-Political Treatise* and the *Political Treatise*, we begin to see the integral nature of Spinoza's project, and the ways in which his metaphysics relates to his politics as well as his ethics. The *Treatises* deal with the conditions under which intellectual felicity would be socially and politically possible, just as the *Ethics* deals with the conditions under which such intellectual felicity would be metaphysically, psychologically, and morally possible. In all these cases, however, there is a pervasive tension, if not an outright contradiction, in Spinoza's formulations of the relation between central authority and individual freedom, between the totality and the individual, between substance and modes.

I suggested at the outset that the contradiction had its sources in the way Spinoza formulated the problem, namely as a problem of constructing the necessary conditions for the satisfaction of the intellect's nature as knowing activity. This calls for an ideal construction: what can we imagine as the optimal conditions for the human mind, i.e. for its activity? We are led through an amazing variety of considerations: not only what God must be like, or what perfection of order and connection Nature must exhibit, so that it can be fully and rationally understood; but also, how the emotions can be understood so that they do not become passions which affect and disturb our lives and distort or inhibit our reason; and further, how our relations with others should be ordered to enhance our own happiness and contribute to theirs. All of this, then, is an attempt to construct a world in thought which can become a guide to our activity, as thinking, feeling, social human beings.

But the tensions are there. Spinoza inherited one world (or two, in fact), and he lived in another (or again, perhaps in two). He inherited, historically, a world whose very form of social organization, and whose conceptual framework focused on the totality. The late medieval world, at least in its ideological and philosophical self-conception, presented itself as an organic community, in which individuals were subsumed in their stations and duties, and where the necessity of God's law was

paralleled by the authority of church, feudal lord and king. The authority of the text (whether the Bible, or the laws) was, if not in practice, then certainly in theory and ideological belief, absolute; and coherence was presupposed, nurtured, and enforced, in interpretations. Spinoza also inherited the world of Sephardic Jewry, recently traumatized by the expulsions from Spain and Portugal, caught up in the tradition of the Book, the authority of the law, and more than usually strictly self-regulating in communal life, in the face of the shifting and hostile politics of the Christian world. The excommunications, first of Uriel DaCosta, then of Spinoza, were symptoms of the fear which division within the community's life, or within its beliefs, engendered. The Jews of Amsterdam were running a tight ship, for many reasons.

On the other hand, Spinoza lived in one of the most advanced commercial centers of its time, where a burgeoning bourgeoisie was contesting older feudal forms, and where the individual's right to pursue the activities which satisfied his desires were being asserted in the marketplace as well as in politics. Hobbesian 'possessive individualism' (to use C. B. McPherson's term) was already fully articulated, and known to Spinoza as a theoretical formulation of the new society, caught as it was between authoritarianism and individual liberty. So, too, Descartes' system began with an 'I', and with the authority of individual reason, and the assertion (though not the practice) of skepticism.

What I am suggesting, in very broad strokes to be sure, is that the particular form of the tension in Spinoza's philosophical system needs to be examined in the context of the historically new tensions between individual and community which were developing in the transition from late feudal society to bourgeois society, in its early mercantilist and capitalistic stages. It seems to me that the contradiction in Spinoza's system between substance and its modes, between the all-determining and infinite totality, and the individual realizing his or her nature in the activity of satisfying desire – that this contradiction has to be understood also in relation to the peculiar position in which Spinoza found himself, at the transition point between two worlds. He tried to hold both world conceptions together. Or rather, he tried to introduce the new world conception into the matrix of the old, and did so with a passion for coherence, and an extraordinary rigor in philosophical construction. But he failed, because he could not succeed. The worlds were incompatible, in thought as in life.

Spinoza's motive, seen in one way, is a very subjective one: the

satisfaction of the intellect's natural desire to know. Why should the world be such that the intellect should be satisfiable in knowing it? Moreover, why should the world (God or Nature) be such that human beings could possibly achieve happiness in it, in pursuing their natural ends? Seen in a different way, Spinoza's motive is a universal and a social one: the system of the world is such that all human beings have not only a capacity for freedom, but the possibility of realizing it through their own activity. The motive, in its historical setting, is a complex one, therefore. It is caught between a desire for order, for peace, for communal wholeness, and a desire for freedom from any imposed authority, for individuality, for the unique and separate striving of the spirit for individual fulfillment, if not for immortality.

The contradiction in Spinoza's theory of individuation has its roots in the conflicts within this motive itself, and, beyond it, in the major conflicts engendered by a great social transition. But this can only be suggested here, and remains the subject for a much fuller and more careful analysis.

NOTES

[1] *Correspondence*, trans. by R. H. M. Elwes Dover Publications, New York 1955, Letter LXXI (LXXXII), pp. 408–9.
[2] *Ibid.*, Letter III, pp. 279–82; and Spinoza's letter to Oldenburg, XV (XXXII), pp. 290–3.
[3] *Ibid.*, Letter XXIX (XII), pp. 317–23.
[4] *Ibid.*, Letters XXXIX, XL (XXXV), XLI (XXXVI), pp. 351–8.
[5] *Ibid.*, Letter L (L), pp. 369–70.
[6] M. Wartofsky, 'Action and Passion: Spinoza's Construction of a Scientific Psychology', in *Spinoza*, (ed. by M. Grene), Anchor Press/Doubleday, Garden City, N.Y. 1973, pp. 329–53. Reprinted in this volume, pp. 231–254.
[7] Abbreviated references in the present paper are to the *Ethics*: 'II P7' stands for *Ethics*, Pt. II, Proposition 7.
[8] *Correspondence*, Letter XXIX (XII), pp. 317–23.
[9] *Ibid.*, pp. 319–20.
[10] *Ibid.*, p. 322.

HUME'S CONCEPT OF IDENTITY AND
THE *PRINCIPIUM INDIVIDUATIONIS*

[1961]

I.

In Hume's account of Identity, an apparent contradicition is presented as the ground for introducing the principle of Imagination as the source of that *fiction* by means of which we attribute not merely continued existence to objects when they are not present to the senses, but an *identity* through time to the objects so conceived. On the grounds of an epistemological critique, the ontological identity of persistent objects is declared fictive. But in the process of this analysis, an alternative ontological scheme is proposed, whose import is often overlooked in accounts of Hume's empiricism. It is obscured by Hume's ambivalent usage of key terms in the *Treatise*, but I think it can be shown to emerge from an analysis of the conceptual elements in his scheme. Such an analysis is instructive not only in terms of what may be inferred from what Hume himself has to say, but also in terms of Hume's dependence on, and transformation of, historically antecedent concept-structures. I hope to show that from Hume's own account, he is to be reckoned not as a clear atomist, but as an ambiguous one, with strong elements of that relationalism which is identified with the tradition of Leibniz and Whitehead. The contemporary relevance of Hume's struggle with the concept of identity is blurred by the shift from Hume's psychologistic framework, to an increasingly logistic and semantic one, in the contemporary discussion of the analytic paradoxes of identity. Whereas Hume talked about 'mental geography,' Ryle talks of 'logical geography'; but the import of a conceptual analysis is that it deals with the structures that are common to any 'geography' of this sort.

The two concepts on which Hume's account of identity may be said to rest are among the most difficult and traditionally problematic in metaphysics: *Time* and *Individuation*. In Hume, they are apparently torn free from the metaphysical context, and are reinterpreted in terms of an apparently simple epistemic scheme. The transformation is fruitful, both

for the history of philosophy and for science; but the problems are not resolved. They reappear in a significantly new way. This attests to the overriding persistence of conceptual structures, through their transformations and evolution, and to the formative significance that they have at any stage and in any particular area of human knowledge.

<div align="center">2.</div>

It would be proper to speak of Hume's *concepts* of Identity. In his analysis, alternative senses of identity are examined, and are reduced either to spurious or empty cases, or to his 'fiction of the imagination.' For purposes of this analysis, four senses of 'identity' may be distinguished, without at this point indicating what relation they bear to each other, or whether in fact they are really different.[1] The following descriptions are not Hume's, though the distinctions are explicitly or implicitly set forth in the *Treatise*.

I. *Logical Identity*

This first sense concerns the uses of ' = ' in logical notation, the questions of synonymy, the substitutivity of terms in propositions, and the issues involved in naming and reference, the problems of using two different names for the same 'thing,' or of using the same term twice in relational or functional statements (e.g. xRx, or (x) F (x,x), for reflexivity), or of the substitutivity of descriptions for names, etc. Apart from adequate analysis of the conventions of notation, and an examination of the consequences, in a logical universe of discourse, of using one rather than another system, there is no 'problem' of identity here at all, in Hume's view, and he dismisses this whole area summarily, as we shall see. The problems arise when the reference of terms, and the significance of operations beyond the universe of logical discourse is considered.[2]

II. *Ontological Identity*

Here, the concern is over the attribution of identity to 'real' entities, whatever the specific ontological scheme is. Whether the entities are mathematical or physical, abstract or concrete, individuals or classes, *sensa* or *qualia* is a matter of indifference at this point. But the distinction between *real* or *ontic* identity, and *virtual*, or *formal*, or

fictive, or *imagined* identity is crucial. Merely nominal identity, then, is clearly excluded, and belongs in I.

III. *Epistemological Identity*

What is at issue, and what clearly distinguishes this sense from the previous two, is its concern with epistemic justification of attributions of identity, or with the epistemic *mechanism* whereby such attributions are made, or with the nature of 'identifying,' within some epistemic framework. The requirement of an account of how identity *is known*, distinct from *what is*, in ontic terms, identical, or from how identity *may be stated*, in ordinary or logical languages, is peculiar to this sense of identity. This, it would appear, is the proper, explicative sense of identity, to which all the others may be reduced, according to Hume. His confusion of senses II and III is significant, as we shall see.

IV. *Teleological or Functional Identity*

Although this is not clearly a 'sense' of identity in the way that I, II, and III are, and becomes for Hume a special case of III, it has a distinct form, and is central in Hume's account of personal identity. This is the identity we attribute to complexes, or aggregates, in view of their organization for some 'common end,' their constitution with respect to performing some function. For Hume, this applies not alone to living organisms or persons, but to any object so constituted (e.g., a ship).

<div align="center">3.</div>

For Hume, the concept of identity is problematical only within the context of his epistemological scheme. It is not identity *as such*, but the inference of identity from variable and interrupted perceptions that appears paradoxical. What identity reduces to, or what it means in an empirical sense, is what concerns Hume. Merely nominal identity, which he considers an empty or spurious case, is summarily dealt with.

... we may observe that the view of any one object is not sufficient to convey the idea of identity. For in that proposition, an object is the same with itself, if the idea express'd by the word, *object*, were in no ways distinguish'd from that meant by *itself;* we really shou'd mean nothing, nor wou'd the proposition contain a predicate and a subject,

which however are imply'd in this affirmation. One single object conveys the idea of unity, not that of identity[3].

If a unity does not convey the idea of identity, then neither does a multiplicity of objects, since "the mind always pronounces the one not to be the other, and considers them as forming two, three, or any determinate number of objects, whose existences are entirely distinct and independent."[4] Yet, we find that the inference we make from the constancy and coherence of successive and distinct impressions is an *individuation*; that is, we

are apt not to regard these interrupted perceptions as different (which they really are) but on the contrary consider them as individually the same, upon account of their resemblance. But as this interruption of their existence is contrary to their perfect identity, and makes us regard the first impression as annihilated, and the second as newly created, we find ourselves somewhat at a loss, and are involv'd in a kind of contradiction. In order to free ourselves from this difficulty, we disguise, as much as possible, the interruption, or rather remove it entirely, by supposing that these inter- rupted perceptions are connected by a real existence, of which we are insensible.[5]

In a brief compass, Hume has opened what we might call a 'stacked' Pandora's Box, filled with monsters of his own creation, among the best known of which may be included *perceptions* (particularly of the interrupted variety), *individuals*, and *sames*. Overwhelmed with these, we will pay relatively little attention to minor demons like *resemblances*, and deal with *real differences* (which 'really are'), later.

4.

A *Unity*, for Hume, as a 'single view of an object' possesses no identity properly speaking. An *Individual* does. Contrary to ordinary usage, names of the sort 'object,' and 'itself,' or ostensive statements like 'O appears' or 'There is an O' do not denote individuals for Hume. A terminological difficulty arises here, related to Hume's concept of 'Individual.' An individual for Hume is not a 'perishing existence' but has duration, exists through time, or, to use Quine's felicitous phrase, is a "time-consûming object."[6] The one-shot terms, 'object' and 'itself,' are merely alternative ostensions for a once-given unity. Insofar as they are not time-referent, they do not extend far enough to encompass an identity. An individual, which is *individuated* precisely in the fact that it can be ostensively referred to at *two or more separate times*, is distinct

from such a unity, which doesn't involve reference to time at all. And once Hume uses 'identity' and 'the Principle of Individuation' as alternative, synonymous expressions, we are, if not in the metaphysical soup, at least in the analytic frying pan.

It becomes paradoxical to talk about a 'single object' except in terms of a single appearance. And the only 'object' which does appear thus singly in Hume's scheme is the impression or perception, which is a unitary 'internal and perishing existence,' and is intrinsically a non-repeatable occurrence. Its only time-reference would appear to be its temporal uniqueness, the fact that it can only 'happen' or 'exist' once; in fact, it is anterior to time, since time, for Hume, is nothing but the 'manner' in which such impressions succeed each other. That we may take this not merely as Hume's epistemological 'simple' but also as his ultimate ontological entity is clear in several passages in the *Treatise*. In the discussion of the ideas of Space and Time, he writes:

'Tis evident that existence in itself belongs only to unity, and is never applicable to number, but on account of the unities of which the number is compos'd. . . . That term of unity is merely a fictitious denomination, which the mind may apply to any quantity of objects it collects together; nor can such an unity any more exist alone than number can, as being in reality a true number. But the unity which can exist alone, and whose existence is necessary to that of all number, is of another kind, and must be perfectly indivisible, and incapable of being resolved into any lesser unity.[7]

If we look past the difficulties concerning Hume's ambivalent usage in different sections of the *Treatise*, to what is the formal characteristic of the impressions or perceptions, we find that they are likewise atomic, irreducible; that is, they share the 'existence' characteristics of 'units.' Furthermore, in regard to their immediacy, and their epistemological indivisibility, Hume writes, as a true heir of Sophistic, Protagorean phenomenalism:

. . . Since all actions and sensations of the mind are known to us by consciousness, they must necessarily appear in every particular what they are, and be what they appear. Every thing that enters the mind, being in *reality* as the perception, 'tis impossible anything shou'd to *feeling* appear different. This were to suppose, that even where we are most intimately conscious, we might be mistaken.[8]

Thus far, we have the unit as the existent *par excellence*, as the ultimate simple of consciousness, and the assertion that such a unit is not sufficient ground for an affirmation of identity, or individuation. There is another model of atomicity which enters into Hume's identity paradox, and that concerns *Time*. Its importance has already been

indicated. When we get to the concept of identity proper, and beyond the spurious case of nominal identity, the distinctive parameter is Time. In the first account of Identity in the *Treatise* (Part I, sect. v). writing of *Relations*, Hume says of identity: "This relation I here consider as apply'd in its strictest sense to constant and unchangeable objects. . . . Of all relations the most universal is that of identity, being common to every being, whose existence has any duration."[9] But Time, or Duration (used interchangeably by Hume; e.g., Part II, sect. iii) is made up of parts which are not coexistent; coexistence belongs to extension, and as such is the origin of our idea of space. The atomicity which is characteristic of existents, and of impressions or perceptions, is also characteristic of time, "which since it appears not as any primary distinct impression, can plainly be nothing but different ideas, or impressions, or objects dispos'd in a certain manner, that is, succeeding each other."[10] Further, "the idea of duration is always deriv'd from a succession of changeable objects, and can never be convey'd to the mind by anything stedfast and unchangeable. For it inevitably follows from thence, that since the idea of duration cannot be deriv'd from such an object, it can never in any propriety or exactness be apply'd to it, nor can anything unchangeable be ever said to have duration."[11] From talking about time as a 'manner' of the disposition of impressions, to talking about 'time as it exists' is, for Hume, merely a matter of turning his attention from the psychological origin of the idea, to the analysis of the idea itself, and no ontological implications about 'existing time' need be derived from this manner of speaking. But it is interesting from another point of view. In isolating the idea of time, Hume writes,

'Tis a property inseparable from time, and which in a manner constitutes its presence, that each of its parts succeeds another, and that none of them, however contiguous, can be coexistent. For the same reason that the year 1737 cannot concur with the present year 1738, every moment must be distinct from, and posterior or antecedent to another. 'Tis certain then, that time, as it exists, must be compos'd of indivisible moments. For if in time we could never arrive at an end of division, and if each moment, as it succeeds another, were not perfectly single and indivisible, there would be an infinite number of coexistent moments, or parts of time: which I believe will be allow'd to be an arrant contradiction.[12]

This atomicity of time, if time is derived from change, makes change completely atomic; there can be no change except as a succession of units. The objective aspect of change is the *nextness* in time (or, more accurately, a nextness which constitutes the idea of time), of particulars, units, or 'bits.' These are not only correlated with the indivisible units of

time, but time is denied any substantive status at all, and is merely the manner of the disposition of these *discreta*. For a particular or unit of this sort (whether it is taken psychologically as an impression or perception, or ontologically as an existent), to *extend over* several units of time would be paradoxical, since the particular, as an impression or an existent, is in effect bounded by, or bounds a time unit. But, Hume points out, it is precisely this paradox that we accept when we have an idea of an unchangeable object. So that the idea of an object persisting in time has absolutely no warrant in the immediate atomicities which constitute the origin of our ideas. What we supply, then, by a 'propensity to feign,' is a 'fiction of the imagination': *Identity* or *Individuation*. We somehow *actively* generate objects, individuals, persistences out of a world which presents nothing but a multiplicity of *discreta*, in one or another relation of nextness.

Since, 'in reality' – i.e., in terms of the ultimate simples which seem to exhaust Hume's ontological commitment – a time-consuming object cannot be a unity, the only other alternative is to consider it as a multiplicity, to conceive of it not as *unity*, but as *number*. But, says Hume, this is clearly not what we mean when we talk of the 'same' object at *different* times. *Sameness*, in individuating objects, is empty if it is not conceived against some background of difference – namely, different times. Real identity, as opposed to mere tautological identity, or merely nominal duplicity, must be constituted by sameness in perception, by the *resemblance* of a past impression to a present impression. And this *resemblance* is, for Hume, as it is later for Mill,[13] an unanalyzable relation, or 'principle of comparison.'[14]

The paradox that Hume himself elicits in his analysis sneaks into his own language again and again, so that he can talk of "a single object, plac'd before us and survey'd for any time, without discovering in it any interruption or variation,"[15] which is the origin of our idea of identity. Now, the time *in which* we survey a single object cannot be derived from the situation of a single object, so that we "imagine . . . a change of time without any variation or interruption in the object." The analytic difficulties increase, in the attempt to 'conceive,' or 'imagine' two points of time 'at the same instant'; or, in the alternative perspective which Hume proposes as a contrast to this, in the attempt to "trace the succession of time by a like succession of ideas, and conceiving first one moment, along with the object then existent, *imagine afterwards* a change in the time without any variation or interruption in the object."

Here then, says Hume, is that "medium 'twixt unity and number; or more properly speaking, *it is either of them, according to the view in which we take it*: and this idea we call that of identity."[16]

<div align="center">5.</div>

There is one more point to be made in exposition of Hume's view which bears on the issue I want to examine. The bastard "mixture 'twixt unity and number," which in fact is no mixture at all, but a perspectival differentiation in Hume, is ascribed to the imagination, as a 'fiction' occasioned by the resemblance of impressions, i.e., their *natural* association engendered by the smooth transition from one to the next. The imagination is propensive in the sense that it is a kind of copulative, or connecting *activity*, an inertial motion, which "when set into any train of thinking, is apt to continue even when its object fails it, and like a galley put into motion by the oars, carries on its course without any new impulse."[17] The *appearance* of continuity, in the atomicity of unit-impressions, is therefore something *added* by this imagining activity. But in Hume, this fiction, which depends on resemblance, involves us in another difficulty. Resembling ideas are "not only related together, but the actions of the mind, which we employ in considering them, are so little different *that we are not able to distinguish them.*"[18] How then are they *resembling ideas* and not merely a 'single view'? Unless *some* distinction can be made, Hume effectively destroys the groundwork here for his account of the fictiveness of identity: i.e., the atomicity, or distinctness, or discreteness of impression-units or time-units. It might be argued then, in Hume's own manner, that if the distinctions are so little different as to be indistinguishable, they are no distinctions at all, and that the fiction consists *in making such distinctions at all.* This is uncomfortably close to the views of Bergson and Whitehead, in which the distinctions within the perceptually given are in fact acts of the intellect or postulates of thought. In Hume's own scheme, the distinction cannot be attributed to sense, or perception, if it is not distinguishable to the senses as such; nor can it be attributed to the imagination which is an active ordering and, it would sometimes appear, a 'gestaltist' faculty.[19] In Hume's scheme there is only one other source for such distinctions – *Reason* – and these must be, then, by default, *distinctions of reason.*[20] At the outset of the *Treatise*, we are told that "simple perceptions, or impressions and ideas are such as admit of no

distinction nor separation.''[21] The correspondence between an impression and 'its' idea is spoken of as an 'exact representation.' *Resembling*, in this sense implies not a distinction between the nature of an impression and its idea, but only a difference in 'degree' between them. It is not in this resemblance between impression and idea, as simples, that any distinction can be made, and the simplicity of these simples, in the sense of their indivisibility, is thus maintained. The *distinction of reason* is introduced with regard to abstract ideas, however. The object of this distinction is not the simple impression or idea itself, but the *relation of resembling*, between different impressions and ideas; i.e., it is a distinction which arises only in the consideration of the relations between resembling perceptions. Hume writes: '' 'Tis certain that the mind would never have dreamed of distinguishing a figure from the body figured, as being in reality neither distinguishable nor different, nor separable; did it not observe that even in this simplicity there might be contained many different resemblances and relations.''[22]

When we make such a distinction in what "formerly seemed, and really is perfectly inseparable" we seem to have separated off *parts* of a *simple*, which would be manifestly contradictory. Hume avoids this by an appeal to an analysis reminiscent of Duns Scotus' doctrine of *Formal Distinction*.[23] The distinctions thus made are not real distinctions, in the sense that real entities are not thus distinguished. They seem instead to be distinctions made in the turning of attention to relations of which such entities are capable: ''. . . we view them in different aspects, according to the resemblances of *which they are susceptible*.''[24] Such a distinction is, then, not merely a virtual or logical distinction. If our ultimate entities in Hume's system are psychological, then the susceptibility of such entities to relational distinctions implies the same kind of double aspect for these simples that the Subtle Doctor had in mind, in his formal distinction: the simplicity is taken care of by the *haecceity*, the 'thisness' of the indivisible existents; their susceptibility to a manifold of resembling relations, by our acts of attention, assumes some *natura communis*.[25] What saves the distinction from being merely logical is precisely the fact that Hume's psychological elements are also his ontological elements. What would appear to be merely a distinction *in intellectu*, as a second intention, becomes *more* than this, but *less* than a distinction *in re*. The unity of our indivisible perceptions is maintained, but the number, of potential relations of resemblance in a unitary perception, is accounted for as well. The distinctions of reason, traditionally problematic in Hume

scholarship, and apparently limited to his treatment of abstract ideas, have however a strange formal or structural similarity to the presentation of the 'fiction of the imagination,' identity: in both cases, a modality is introduced between the apparently exclusive disjuncts, Unity and Number. We might characterize this as a distinction between simplicity in existence, and multiplicity in relations. We can, therefore, make relational distinctions within what is, in existence, simple and indivisible, or more accurately, what is, in perception, simple and indivisible; every perception being already involved in an ontological presumption: "Whatever we conceive, we conceive to be existent. Any idea we please to form is the idea of a being; and any idea of a being is any idea we please to form."[26]

Are we offered here any way of dealing with Hume's paradoxical notion of "indistinguishable differences"? Hume himself is only occasionally aware that he is using this notion paradoxically, but he is constantly aware of the difficulty of the problems he is up against. Witness the repeated apologies with which he prepares us for "abstruse reasoning," his warning that he will have to have recourse to "profound metaphysics," etc. Although I don't think that Hume pretends to have a solution, and although he explicitly pretends to be satisfied that a certain dissatisfaction is the only conclusion which avoids either the 'vulgar view' or the alternative philosophic monstrosities, there is a warrant in Hume for what would appear at the outset as an unHumean conclusion. Hume's suggestion concerns *the relativity of difference*.

6.

In attempting to avoid the paradox of indistinguishable differences, in a spare phenomenalistic conceptual scheme, we may assume thresholds of perceived difference. There is a warrant for this in Hume. As indicated above, the perception of change is, for Hume, a perception of a difference in perceptions, i.e., the awareness of the atomicity of impressions. But what determines this awareness of the distinction between one unit and another? It is, according to Hume, not an absolute difference, but a relative or proportional degree of difference. In some sections of the *Treatise*, it would appear that the distinction or separability of simple impressions or ideas is merely a matter of simple inspection, uncomplicated by relations. Hume catches us in a circularity, early in the *Treatise*, when he asserts that whatever is different is

distinguishable, and that whatever is distinguishable is separable by thought and imagination. But, conversely, he adds that whatever is distinguishable must therefore be different, "for how is it possible we can separate what is not distinguishable, or distinguish what is not different?"[27] But the 'distinguishables' here are not atomic perceptions as such; they are *degrees* of difference in quantity or quality. This is problematic, because what is assumed is *a degree of difference in the same perception*, and therefore some change or alteration *in* it, without reference to anything else. This would involve us in the same paradox that the identity of objects involves us in; or in a *petitio principii* in which the distinction of a difference of degree between *two* perceptions presupposes that they are distinctly *two*, whereas the only ground for this latter distinction *is* the distinction of a degree of difference.

On the one hand, change is taken atomistically – is made manifest phenomenally only by radical discontinuities, the occurrence of new, different, unrepeatable perceptions. The perception is not infinitely divisible, but is indivisible, a discrete quantum: in principle, therefore, *absolutely* discrete, separate in 'reality.' Additional problems are raised when we consider that as a time unit, such a perception is not presented (or represented) as instantaneous, but has a finite duration, as when Hume speaks of impressions as "internal and perishing existences" succeeding each other with an "inconceivable rapidity." Yet, they cannot have even this problematic duration, or present a 'piece' of time, since the passage of a finite duration, if *not* instantaneous, would require a 'marking off' of some time, i.e., some time metric, which by definition is impossible to establish on any other grounds but that of a succession of perceptions. In still another version of this same quandary, Hume's account of the smooth transition between resembling impressions makes the implicit assumption that a change, or difference *actually* occurs, but disturbs us so little that we don't *notice* it. The distinction remains then, a distinction by postulate rather than an empirical one, and yet the whole empirical scheme is based on the initial assumption of such distinctions. Despite these difficulties, Hume is withal more of an empiricist than such an account of postulated difference, at the base of his conceptual framework, would make him appear. In a quasi-operational and pragmatic discussion, in the appendix to his section on the concept of a vacuum, he writes:

As long as we confine our speculation to *the appearances* of objects to our senses, without entering into disquisitions concerning their real nature and operations, we are safe

from all difficulties, and can never be embarrass'd by any question. Thus, if it be ask'd, if the invisible and intangible distance, interpos'd betwixt two objects, be something or nothing: 'Tis easy to answer, that it is *something*, viz. a property of the objects, which affect the *senses* after such a particular manner. If it be ask'd, whether two objects, having such a distance betwixt them, touch or not: It may be answer'd, that this depends upon the definition of the word, *touch*. If objects be said to touch, when there is nothing *sensible* interpos'd betwixt them, these objects touch: If objects be said to touch, when their images strike contiguous parts of the eye, and when the hand *feels* both objects successively without any interpos'd motion, these objects do not touch. The appearances of objects to our senses are all consistent; and no difficulties can ever arise, but from the obscurity of the terms we make use of.[28]

By analogy to this reasoning, reminiscent of William James' squirrel story, it would seem that difference depends on *perceptible* difference. Indeed, if the changes in an object are imperceptible, then in fact there are no changes, and identity reduces to unity. That to which we ascribe a perfect identity, because of the imperceptibility of changes in it, cannot be known to perception as changing, and would therefore, on Hume's own grounds, *be* a perfect identity. (Leibniz expresses this strikingly: "God will never choose between indiscernibles."[29]) If we choose, with Hume, to postulate that the condition for perfect identity is invariance and unchangingness, then the paradox is set: only a simple is invariant and unchanging, and can have no reference to time, since time derives from a succession, i.e., a requirement of more than *one*. Hume's account of the natural propensity to feign identity is based on the assumption that *on inspection*, every such identity turns out to be *imperfect*; that what appears the same is in fact different. But it would seem that unless the operation of differentiating can be reasonably accounted for *in the first place*, the positive activity of the imagination in feigning sameness with respect to difference would appear not 'natural,' but gratuitous. The paradox of identity would vanish, unless one of its contradictory conditions is as well accounted for as the other, and within the same conceptual scheme, i.e., phenomenalism. The way to clear up a paradox is, of course, to point out that what is contradictory about the *same* thing is not contradictory if we are talking about two different things. We might, following Hume's operational definitions of 'touching' (see above) end up with a satisfactory formula, such as Quine's:

"Objects indistinguishable from each other within the terms of a given discourse should be construed as identical for that discourse."[30]

If we take Hume's apparently pejorative use of 'fiction' and 'illusion' in a more liberal sense – in one, by the way, for which there is warrant in

Hume[31] – then this is precisely what Hume proposes. No man more ready
to go along with the *consensus gentium* than he, when his universe
of discourse is the tavern or a game of backgammon. The 'naturalness'
of the propensity to feign, and even the historical antecedents of his use
of 'fiction,' in the Scholastics, in Malebranche, in Locke, and his stated
intention to develop a science of Man, a natural philosophy of Mind all
indicate that 'feigning' is a proper and natural operation in Human
Nature. But before we are carried away by a tolerant good will, Reason
enters, and creates intolerable conditions – and we are caught between
the 'vulgar' and the 'philosophical' universes of discourse, with no easy
mediation, nor even the comfort of a conventional agreement to talk out
of both sides of our mouths. The solution to the paradox, in terms of
pointing out that it is two different things we are talking about, has been
proposed by critics of Hume from the days of Green and Grose on.
Marhenke proposes such a solution, and discovers the source of the
confusion in Hume's inability to make a distinction between objects and
events, and between change and succession. He writes:

Hume's denial that an unchanging object is temporal is traceable to his failure to
distinguish between objects and events. In his philosophy impressions play the role of
objects and also of their states. Now, when we are aware of a changing object, we are
aware of a succession of states which are qualitatively different. Similarly, when we
observe an unchanging object, we are aware of a succession of states which are qualita-
tively identical. A succession of states can of course be apprehended without our appre-
hending it as either change or invariability. In order to apprehend the succession as
either one or the other, however, it is necessary that we be simultaneously aware of the
perdurance of the object.[32]

The distinction between objects and events may clear up the paradox,
but the price is too high: Marhenke and others who propose this sort of
solution are ready to abandon the possibilities of a solution in
phenomenal terms before the possibilities have been exhausted. The
readiness to accept an alternative realist conceptual scheme, as offering a
solution, is based in part on the inability of Hume's phenomenalism to
deal adequately with the problem, an inability made manifest by the
appeal to a *Deus ex Machina*, a creative gestalt-forming imagination. But
I think that a closer examination of Hume offers a lead to a phenomenal
solution. And this is in his aforementioned notion that perception of
change is a matter of degree of difference, i.e., that perception of change
is relational. But the perception of change lies not only at the basis of the
problem of connecting what is *given as distinct*; but is also at the basis of

the problem of establishing the *distinctness of the given*. In a strictly phenomenal scheme we can't merely postulate that what is different is distinguishable, but must ultimately base difference on distinguishability; in which case, what is required is a phenomenal account of the ability to distinguish difference. In such a scheme, the perception of change is a perception of the distinction between one perception and another – the establishment, *in the first place,* and on epistemic grounds, of the very atomicity which Hume assumes as given.

<div align="center">7.</div>

In discussing the gradualness of some of our perceptions of change, Hume offers a notion of thresholds of perceptible change, in terms of the relativity of our perception of difference. This is developed in his discussion of personal identity, but comes up as well in his discussion of the concepts of a mathematical point and of a vacuum, of abstraction and the distinction of reason, and of the relation of resembling. In the discussion of personal identity, Hume extends the concept of identity in two ways. First, Hume introduces the notion of a constitutive identity in which parts may *change place, move,* but in which nothing is perceptibly *added* or *taken away.* In other words, he introduces a notion of extensive or quantitative identity. Second, a notion of functional identity is introduced, in which there is a relation of parts to some common end, and further, there is a 'sympathy' of parts in relation to a common end. In respect to the first, he writes:

> . . . tho' the change of any considerable part in a mass destroys the identity of the whole, yet we must measure the greatness of the part, not absolutely, but by its proposition to the whole. The addition or diminution of a mountain wou'd not be sufficient to produce a diversity in a planet; tho' the change of a very few inches wou'd be able to destroy the identity of some bodies.[33]

The notion of *Same* and *Diverse* reduces, phenomenally, to the perception of an interruption in what was previously uninterrupted and unchanging in perception. But the threshold of the perceptibility of such an interruption depends not on some absolute unit of change; on the contrary, the unit itself depends, for its distinction, on a proportional difference from some whole, i.e., to a relational difference against some background. But let us extend this reasoning beyond the *objects* which are said either to change or to remain invariant, and apply it to our *original* units, i.e., the impressions, or perceptions. We get then to the

heart of Hume's presupposition, and by applying Hume to Hume, come up with an inversion of Hume's two principles, which, in the tortuous re-examination in the *Appendix*, he finds he can neither render consistent, nor renounce: "viz. *that all our distinct perceptions are distinct existences, and that the mind never perceives any real connection between distinct existences*."[34] The inversion asks instead how a perception is known to be distinct, and how the mind perceives any real distinction between perceptions, or existences; by what phenomenal means do we establish discriminations within the given. This is as real a question in Hume as the first, and although the whole structure of the *Treatise* is built on an initial presupposition of the distinctness of perceptions, yet Hume himself, by backing himself into it, offers a clue to an analysis of the presupposition itself. If we apply Hume's criterion of difference to his 'atoms,' then these atoms themselves are relationally determined as to their very atomicity. To carry this further, 'an' impression is simple, only to the extent that it is *proportionately differentiable* from some other impression. 'A' time unit is simple, and indivisible, on phenomenal grounds, only by virtue of the fact that its initial and terminal boundary is perceptibly interrupted. There are no other grounds for the distinction in phenomenal terms. What this would seem to indicate is that what Hume considered in another connection as *distinctions of reason* turn out to be the only grounds on which we can establish the distinctness of those very simples whose *susceptibility to a manifold of relations* permits us to make distinctions *within* them. Here, however, it is their susceptibility to a manifold of phenomenal relations that permits us, in the first place, to make distinctions *of* them.[35] Within such a phenomenal framework, it is possible to talk about the *relative* persistence of an 'object' or of an 'impression,' indifferently, only if we have, in any given view, *both* a succession and an invariance. The notion of *proportionate* change involves a *relation of difference*, as much as it involves a *relation of identity*. Our *discreta* cannot be known as such *apart from* a relative invariance, just as an invariance cannot be known apart from a relation to relatively discrete appearances. Our 'simples' are no more *given* phenomenally then the persistent, or time-consuming objects which they allegedly fictively constitute. Apart from one such relation of proportionate difference, which thereafter becomes a subject for experimental, empirical investigation, the sense in which Hume uses 'same,' 'unchanging,' and 'uninterrupted' remains either completely opaque, or crudely intuitive, or unphenomenally postulational.

8.

The outcome of all this is that the very units with which Hume starts require some relational distinction to be conceived of as units in the first place. Neither identity nor difference can be conceived *intrinsically* and if Hume is best known for establishing this with regard to *identity*, it should not be overlooked that a consistent phenomenalism requires the same qualification with regard to *difference*. If even the unit, the unit-impression, or the unit-time, is a relational unit, i.e., is 'unitary' in virtue of its relational separability, or distinctness, then it cannot be in any sense an independent starting point, as Hume assumes it to be in much of his discussion. I assume this is what Bradley had in mind when he remarked that "the English writers who have objected to identity have left their principle of atomism and their principle of relativity simply standing side by side."[36] Whether we leave them side by side, or go down the primrose path with Bradley and establish them as internally – I sometimes suspect *infernally* – related, it is certain that it is only by conjoining both somehow that the alternative arbitrary postulates:—(1) of a blocky 'whole' (or a fudgy totality, to take a softer view) or (2) of an ultimate simple, caught in a linear infinity of similar but radically discontinuous simples – that this either/or can be avoided. Whitehead is more generous to Hume, and credits him with explicit recognition of the issue: "Hume discovered that an actual entity is at once a process and is atomic."[37] With this view of the relationality, not alone of identity but of difference as well, Hume's concept of an *individual* supplants his concept of a *particular*. The traditional distinction between *numerical* and *specific* identity becomes harder to maintain, in any phenomenal sense. Both *unity* and *number* then, in terms of Hume's own usage, can be conceived of as 'fictions,' i.e., as dependent concepts. An *individual* is understood by Hume to be *individuated* in the operation of establishing relations (resembling, etc.), and is therefore not a simple but a complex. If our 'distinct perceptions,' which are at the same time 'distinct existences,' share in this characterization of individuality, then it is possible to conceive of Hume's *Principium Individuationis* as more than a fiction of the imagination necessitated by the paradox of identity. It becomes rather the primary mode of conception of both discreteness and continuity, of variation and invariance, within a phenomenalistic framework. On such a view, Hume, albeit in a halting and unfulfilled way, is seen as much closer to the tradition of Leibniz and Whitehead

than is ordinarily assumed. This is not to say that Hume is not an exemplar of the atomistic tradition in British empiricism. But this view has tended to obscure Hume's inheritance not only of the problems, but of the terminology and conceptual framework of a tradition often presented as radically distinct from his. This is no more than to say that atomism is not cut from whole cloth, and that conventional labels often obscure the complexities, the variety, and the alternative possibilities of interpretation and development in any dominant philosophical tradition.

NOTES

¹ Cf. Gustav Bergmann's discussion, in his critical review of Strawson's *'Individuals: An Essay in Descriptive Metaphysics'*. Bergmann writes:
"I distinguish three uses of 'same'; call them logical, epistemological, and ontological. 'Red or blue' and 'not-(not-red and not-blue),' are analytically equivalent. Their referent(s) are not two but one (the same). This illustrates the logical use. For my point it does not matter; I mention it merely to avoid misunderstandings. Whether or not 'two' things *are* really one (the same), is one thing. Whether or not we *know* it is another thing. In the first context (they *are*), 'same' is used ontologically; in the second (we *know*), epistemologically. To emphasize this distinction, as I do and Strawson doesn't is to perform one of those flanking maneuvers the epistemological predicament forces upon us." 'Strawson's Ontology,' *Journal of Philosophy*, **LVII**, 19 (p. 605, Sept. 15, 1960).
² See, for example, the extensive discussion of the analytic paradoxes of identity, in A. Pap, *Semantics and Necessary Truth* (Yale, New Haven, 1958), Chpts. 9–10, esp. pp. 275 ff.
³ David Hume, *A Treatise of Human Nature*, (ed. by L. A. Selby-Bigge) Clarendon Press, Oxford, 1888. p. 200. (All subsequent references are to this edition, and will be footnoted S.B.)
W. V. Quine, in a discussion of this passage in *Word and Object* (Technology Press, Cambridge, Massachusetts, 1960), pp. 116 ff., attributes Hume's view on this to a confusion between sign and object. Quine writes: "If identity is taken strictly as the relation that every entity bears to itself only, [Hume] is at a loss to see what is relational about it and how it differs from the mere attribute of existing. Now the root of this trouble is confusion of sign and object. What makes identity a relation and '=' a relative term, is that '=' goes between distinct and occurrences of singular terms, same or distinct, and not that it relates distinct objects." Quine attributes a similar confusion to Leibniz' explanation of identity "as a relation between signs, rather than between the named object and itself," and sees this confusion in one form or another, in Frege, Korzybski, Whitehead, and Wittgenstein. He cites Wittgenstein's distinctly Humean formulation in the *Tractatus* ... (5.5303): ". . . to say of *two* things that they are identical is nonsense, and to say of *one* thing that it is identical with itself is to say nothing." The argument is based on an interpretation of signs and reference. But Hume's rejection of nominal identity as empty is based on an ontological commitment to *perceptions*, as the only admissible entities in his scheme. On these grounds, identity is an *ontological*

relation between *these* existents. To name *twice* for Hume, would therefore be to have
two separate occasions, two distinct *perceptions* to be named. Anything less cannot
constitute a relation, within his phenomenalistic ontology. Synonymy may be relational
in another universe of discourse, but strictly speaking, for Hume there is only one such
universe. All others are fictive or illusory. Cf. also A. Pap. *Op. Cit.*, p. 276: "If we
replace the ontological mode of speech, 'X and Y are identical concepts (or attributes)' by
the semantic mode of speech ' "X" and "Y" are synonymous expressions,' we only
reformulate, we do not solve the problem. "Hume does essentially the same kind of
reformulation, in replacing the more traditional realist ontological approach to
identity with a phenomenalist epistemic approach, the consequences of which we will
examine here. See also, for an important recent discussion, the Symposium on the
Principle of Individuation, J. Łukasiewicz, E. Anscombe and K. Popper, *Proceedings
of the Aristotelian Society*, Supplementary Volume **XXVII**, (1953) pp. 69–120.

⁴ S.B., p. 200.

⁵ *Ibid.*, p. 199.

⁶ W. V. Quine, 'Identity, Ostension and Hypostasis,' *From a Logical Point of View*
Harvard Univ. Press, Cambridge, Massachusetts, 1953, p. 66.

⁷ S.B., pp. 30–31. The use of a singular term for a collection of perceptibly different
objects is a 'fictitious denomination' in somewhat the same way as the attribution of
identity to a succession of perceptibly different impressions or ideas is a 'fiction of the
imagination.' That the fiction may have its origin in a linguistic economy is suggested
by Hume, but it is Thomas Reid who grasps this more clearly, in his discussion of
identity:

"It may be observed that the identity of the object of sense is never perfect. All bodies,
as they consist of innumerable parts that may be disjoined from them by a great variety
of causes, are subject to continual changes of their substances, increasing, diminishing,
changing insensibly. When such alterations are gradual, *because language could not
afford a different name for every different state of such a changeable being, it retains the
same name and is considered the same thing. . . .* The identity which we ascribe to bodies
. . . is not a perfect identity; it is something which, *for the conveniency of speech, we call
identity.*" (All emphases mine – M.W.)

Essay III, 'Concerning Memory,' *Essays on the Intellectual Powers of Man*, in *Works . . .*
(Duyckinck, Collins and Hannay, New York; 1822), II, pp. 138–9. Cf. with Quine's
similar 'Gavagai' discussion in *Word and Object*, pp. 52–4, and with Quine's claim for
the advantage of formal simplicity gained by the 'central function' of the concept of
identity "in specifying spatio-temporally broad objects by ostension" (*From a Logical
Point of View*, p. 70). The traditional nominalist critique has been that this ostensive
procedure or economy of referring leads to a hypostatization of the referring term. In a
criticism of Hume's nominalist objections, T. Penelhum proposes an apparently simple
solution ('Hume on Personal Identity,' *Philosophical Review*, pp. 571–89, Oct. 1955) in
terms of the 'singularity' of a class name, Φ naming a class of successive objects, A,B,C,
D,E,F, etc. Penelhum appeals to ordinary usage (pp. 579–81) but he misses Hume's
argument on abstraction, and also the import of the phenomenal framework, in
which a class name is nothing but a 'mark' in the Hobbesian and Berkeleian (ultimately,
in the Ockhamist) sense. The problem in Hume is not analytic, but ontological, since
it is the perceptions *as existents* which are discrete and successive, whatever we 'name'

a group of them. On Penelhum's view, class names replace classes named, as *elements* in perception. Hume's nominalism precludes this.

[8] S.B., p. 190.

[9] *Ibid.*, p. 14.

[10] *Ibid.*, p. 37.

[11] *Ibid.*, p. 37. Cf. Whitehead: "Time is known to me as an abstraction from the passage of events." *The Concept of Nature*, Michigan Univ. Press, Ann Arbor, 1957, p. 34.

[12] *Ibid.*, p. 31.

[13] J. S. Mill, *A System of Logic* (Harper, New York; 1879), Bk. I, Chpt. iii, sect. 11, pp. 60 ff., and Chpt. v, sect. 6, p. 83.

[14] S.B., p. 14.

[15] *Ibid.*, p. 201.

[16] (My stresses – M.W.) This condition of viewing the 'same' object at 'two' times would therefore require a recognition, or a *re*perception, impossible *ex hypothesi* in Hume. Cf. Whitehead, *op. cit.*, p. 143: "Objects are elements in nature which do not pass. The awareness of an object as some factor not sharing in the passage of nature is what I call 'recognition'. It is impossible to recognize an event, because an event is essentially distinct from every other event. Recognition is an awareness of sameness. . . . I use recognition for the non-intellectual relation of sense awareness which connects the mind with a factor of nature without passage." Whitehead's 'events' are Hume's 'perceptions,' but as these are the ultimate existents in Hume's scheme, the 'objects' as 'same' are reduced to fictions. On the use of the 'object-event' distinction to criticize Hume, see below, Marhenke, Note 32.

[17] S.B., p. 198.

[18] *Ibid.*, p. 61 (My stress – M.W.)

[19] The imagination is capable of some kinds of distinction, however, and this may be confusing. Early in the *Treatise* . . . Hume writes "Wherever the imagination perceives a difference among ideas, it can easily produce a separation" (p. 10). What imagination does, it appears, is to separate *complex* ideas and reshuffle their elements: "The liberty of the imagination to transpose and change its ideas," in which it differs from memory, which "is ty'd down in that respect, without any power of variation." But in terms of simple ideas, derived from *indivisible* impressions, no internal 'separation' can be made by the imagination, dependent as it is on differences among the impressions, i.e., on *already distinct* impressions, among which, and not within which, it can 'perceive' differences. It doesn't, therefore, make distinctions, but avails itself of them, for creative re-ordering.

[20] In relation to this, there is a cogent analysis of Husserl's critique of Hume's distinctions of reason in the *Logische Untersuchungen*, and a criticism of N. Kemp Smith's position on this, in Robert E. Butts, 'Husserl's Critique of Hume's Notion of *Distinctions of Reason*,' *Philosophy and Phenomenological Research*, **XX**, 2 Dec. 1959, pp. 213–221.

[21] S.B., p. 2.

[22] *Ibid.*, p. 25.

[23] See C. R. S. Harris, *Duns Scotus* Clarendon, Oxford, 1927, **II**, pp. 95, 114–15; and also M. J. Grajewski, O.F.M., *The Formal Distinction in Duns Scotus* Catholic University of America, Washington, D.C., 1944, pp. 35 ff. I am indebted to my colleague, Dr. Erazim Kohak, for discussion and suggestions on this point.

²⁴ S.B., p. 25. See also N. Kemp Smith's discussion, in *The Philosophy of David Hume*
Macmillan, London, 1949, pp. 264 ff.

²⁵ Cf. M. J. Grajewski, *op. cit.*, p. 143. The susceptibility of the *indivisible* impressions
or simple ideas to distinctions of this sort, is not then merely an act of thought, but is,
if our impressions are our ontological simples, an ontological distinction *within* those
simples, and yet less than a distinction *between* these simple existences. Herein lies the
similarity to Duns Scotus' formal distinction. C. R. S. Harris writes (*op. cit.*, pp. 114–
115), regarding Scotus' formal distinction:

"Formal distinctions they are, but formal *a parte rei;* their ground lies in the objective
nature of the reality we are considering, quite independently of the peculiar nature of
our subjective thought processes which lead us to make them. There is thus a real
factual difference *in re*, which corresponds to our distinctions, but it is smaller than a
distinctio realis, that is, a distinction between one *res* and another *res*. For example, in
Socrates, the intellectual and sensitive soul may be conceptually distinguished, though
they are not two things, for Socrates has only one soul. And yet the distinction is not a
mere fiction of the mind, for Socrates' sensational and intellectual processes are objec-
tively different."

²⁶ S.B., p. 67.

²⁷ *Ibid.*, p. 18.

²⁸ *Ibid.*, pp. 638–639.

²⁹ G. W. Leibniz, *Philosophical Works*, (ed. by G. M. Duncan) Tuttle, Morehouse &
Taylor, New Haven, 1890, p. 273; quoted in B. Russell, *The Philosophy of Leibniz*
Geo. Allen & Unwin, London, 1900, pp. 220.

³⁰ W. V. Quine, *From a Logical Point of View*, p. 71.

³¹ See N. Kemp Smith's argument on this, *op. cit.*, pp. 133–37.

³² Paul Marhenke, 'Hume's View of Identity,' *University of California Publications in
Philosophy*, XX, (1937) p. 174.

³³ S.B., p. 256.

³⁴ *Ibid.*, p. 636.

³⁵ Whitehead's discussion of Hume's 'simple' is relevant here. See *Process and Reality*
Macmillan, London, 1929, p. 202:

"We may doubt whether 'simplicity' is ever more than a relative term, having regard to
some definite procedure of analysis." ˴
Cf. his discussion on the "indeterminateness of sense awareness" in *Concept of Nature*,
p. 59.

³⁶ F. H. Bradley, *Appearance and Reality*, Rev. ed.; Clarendon, Oxford, 1930, p. 309,
fn. See also his discussion on 'sameness and difference,' p. 308.

³⁷ A. N. Whitehead, *Process and Reality*, p. 213.

DIDEROT AND THE DEVELOPMENT
OF MATERIALIST MONISM

[1953]

In the much maligned and neglected philosophy of eighteenth-century France, Diderot stands out as the leading, most creative, and revolutionary force. This essay, reworked from a longer unpublished work, attempts to indicate some of the roots of Diderot's materialism and also the developments which his own thinking produced. Acknowledging his debt to Descartes and to the mechanist tradition in France and England, we are here most concerned with Diderot's struggle to break through the limits of this reductionist mechanism, to cope with the contradiction in the materialism of the mechanists, which gave rise to the idealist critique of mechanism and to the idealist solutions of vitalism and hylozoism. Therefore we deal here with Diderot's relation to Spinoza and Leibniz, and to Robinet, Maupertuis, and La Mettrie. The logic of change, motion, transformation, which had been developed in the main by the idealist philosophers and whose absence was the Achilles' Heel of mechanism, is grasped by Diderot not in its idealist form, but rather, as the logic of a material universe, which itself changes, moves, is in constant transformation. In a sense, Diderot puts the idealistic logic of change back on its feet, starting from the outermost limits of mechanism. Taking off from the opposite pole of idealist metaphysics and Hegelian dialectics, Feuerbach, and later Marx and Engels, develop a materialist philosophy of change, of constant transformation, which seems closely akin to Diderot's thought.

Diderot's place in the history of philosophy too often suffers the common fate of the allegedly shallow philosophical writings of eighteenth-century France. It is sandwiched in as a necessary evil between Kant and Hegel, or between Descartes and Bergson, or as a weak continental reflection of British mechanism and empiricism. Past that it is forever consigned to courses in French literature, where its struggle for philosophic recognition can only be a limited one. It is our hope that the renewed interest in Diderot will serve to bring his philosophy forward, not merely as another side of a prolific and creative genius, but as an important contribution to the development of philosophic thought.

THE UNITY OF THE MATERIAL WORLD: DIDEROT AND SPINOZIST MONISM

The universe, for Diderot, was a great chain of being, in the Newtonian tradition. It was an uninterrupted chain, each of whose parts was necessary, the whole having its law of motion and development. Unlike the chain of being of Newtonian mechanism, however, the laws of motion and change were not formal laws set down by the Prime Mover of this perpetual motion world-machine. These laws of motion and change were for Diderot inherent in matter itself. There is nothing outside of the one substance of the universe, no creator, no God: "The supposition of any being whatever placed outside of the material universe is impossible. Such suppositions must never be made, for nothing can be inferred from the" (A.-T., II, 69).[1] And again: "There is no more than one substance in the universe, in man, in animal" (A.-T., II, 117).

This cosmological monism is here based on two premises. First, the supposition of any being outside of the material universe is useless. From an Occamite point of view, Diderot is excluding hypotheses that are not necessary or useful. Nothing can be inferred from these hypotheses, and by the Law of Parsimony, the logical consistency of this monism excludes all immaterial being, i.e., being outside the *material* universe. The logical consistency set up here as a criterion is essentially a utilitarian, practical one, but is rooted not in logical consistency alone. If it were, Diderot could have stated tautologically that there is no being outside of the universe, outside of the one substance of the universe. This latter is the metaphysical monism of Spinoza. Substance is regarded by Spinoza as all that is. God is; God is substance. Immateriality is; immateriality is substance. Spinozistic substantive monism is indeed the logic of Diderot's monism, but the "one substance" of Diderot is not substance in the old sense. It is no longer the ground, the support, the infinite unqualitied subject of the inherence of thought and extension. In Diderot, substance as a category becomes materialized. The one universe is universally material. The logical monism has ontological grounds with which it is inextricably bound.[2] The universe is monistic only insofar as *matter* is the unity of the universe.

The materialist aspect of Occam's razor comes into significant play here, because it is exactly those phenomena which have been attributed to divine, magical, and miraculous agents, ideal and immaterial, that Diderot deals with in terms of material causation. All divine and

miraculous cosmologies fall under this razor. They are reduced to arguments from ignorance. The deist God, too, lost and expatriate somewhere beyond the mechanically functioning world-machine, is denied even this humble place. The blind mathematician Saunderson states on his death-bed that what cannot be explained by natural causes remains a problem to be explained only by natural causes. There are no problems that can be explained otherwise:

Couldn't we have a little less pride and a little more philosophy in our discussion? If nature offers us a difficult knot to untie, let us leave it for what it is, and let us not, in order to untie it, use the hand of a being who will then become himself a new knot for us, more irresoluble than the first. . . . My friend, confess your ignorance and don't bother me with your elephant and your tortoise . . . (A.-T., I, 308).

The elephant and the tortoise of Indian cosmology are equivalent for Diderot to all arguments for God, or supermundane intelligence, from design, arising from an inability to explain natural phenomena materially. The logic of this materialist cosmology is, then, a scientific logic. Moreover, limits of science, or of the knowability of the universe scientifically, do not exist. Diderot's universe is fully knowable; there are no secret, hidden realms. Furthermore it is fully knowable as a material universe. This full-fledged, frank, materialist monism requires that he explain what is meant by *matter*. The substitution of the word matter for the metaphysical concept of substance is not, in itself, materialism. What is understood by Diderot as this matter which is the universe? A fuller elaboration of this will be given later; however, cosmologically, matter, like substance in Spinoza, is the cause of itself, is uncreated. But it is not a substratum, a ground. If it were, Diderot would be up against the dilemma of substance-epistemologies; only the phenomena, the surface, sensory aspects, and appearances of matter would be knowable. But Diderot has in effect stated that there can be no such substance, no problem of a substance, absolutely for itself, a thing-in-itself. If there were, then it would become the elephant or the tortoise, the refuge for ignorance. Spinoza offers the metaphysical basis for a monist solution of this problem in terms of substance.

By virtue of the identity that Spinoza establishes between substance and its attributes and modes, thought, knowing, as one of the attributes of substance, is infinite and eternal (*Ethics*, I, xix); i.e., the attributes of substance *are* the very way in which substance exists, the way substance *is*. This identity of substance and its attributes, translated by Diderot into

the assertion of the knowability of a material universe would mean that knower and knowable, being only in a material universe, are both material: this is the identity between them. Therefore a substance outside of this identity cannot exist; a thing-in-itself, a substratum of phenomena shut off absolutely from knowability, cannot exist. The precondition of the knowability of the world is for Diderot the unity of all being, its materiality. All philosophy is founded on this unity: "If phenomena are not linked one with another, then there is no philosophy" (A.-T., II, 57).

As far as the appearances or phenomena of matter are concerned, Diderot assumes that with the development of science they will increasingly be shown in one general physical theory of the material universe, linked to the essential attributes of matter. For example, he already speculates that electricity and magnetism are two forms of the same essential attribute:

In nature, we shall see, when experimental physics is more advanced, that all phenomena, whether of weight, or elasticity, or of attraction, or of magnetism, or of electricity, are no more than different faces of the same affection. But how many intermediate phenomena are there between known phenomena that are related to one of these causes, that still have to be found in order to make this liaison, to fill in the gaps and to demonstrate this identity? That is something that cannot be determined. [A.-T., II, 42].

The appearances or phenomena are for Spinoza, as for Diderot, merely faces of the same essence:

Matter is everywhere the same, and . . . except insofar as we regard it as affected in different ways, parts are not distinguished in it; that is to say, they are distinguished with regard to the mode, but not with regard to reality. For example, we conceive water as being divided, insofar as it is water, and that its parts are separated from one another; but insofar as it is corporeal substance, we cannot thus conceive it, for as such it is neither separated nor divided. Moreover, water, insofar as it is water, is originated and destroyed; but insofar as it is substance, it is neither originated nor destroyed. . . . I do not know why matter should be unworthy of the divine nature. . . . [*Ethics* I, xv, Schol.]

The language is similar too: Spinoza, in the Proposition in which he proves the divinity, i.e. the eternity and infinity of matter, speaks of it as an attribute of God, or an affection of an attribute. "Different faces of the same affection" in Diderot is the same statement as Spinoza's. This scientific intuition of a General Theory of the material universe is based on the assumption of the knowability of this material universe; furthermore, it is based on a correspondence between scientific theory and the unity of the material universe. Spinoza says: "The order and connection of ideas is the same as the order and connection of things"

(*Ethics* II, vii). It is because the material universe is a unity, is monistic, that our science will ultimately lead us to a physical theory of this unity, to "demonstrate this identity," says Diderot in effect.

Thus, the philosophic monism of Diderot's cosmology is not only a theory of cosmology, but a theory of the correspondence of philosophy and science with the material universe. However he does not point out the accomplishment of this theory as the end of all science. The number of 'intermediate phenomena' is infinite. It is not that Diderot cannot determine how many there are, but that this number 'cannot be determined.'

In this materialist monism of Diderot, the common materiality of body and mind is given as the answer to Cartesian dualism. Spinoza had answered that there can be no two substances (*Ethics* I, x, Schol.; I, xv. Corr. 1). For Spinoza, extension and thought both inhere in one substance (*Ethics*, I, i – ii):

Each attribute of substance must be conceived through itself . . . and . . . although two attributes may be conceived as really distinct, . . . we cannot nevertheless thence conclude that they constitute two things, or two different substances; for this is the nature of substance, that each of its attributes is conceived through itself, since all the attributes which substance possesses were always in it together, nor could one be produced by another; but each expresses the reality or being of substance (*Ethics*, 1, x and Schol.).

A real distinction between the two attributes, thought and extension, exists; nevertheless, the category of substance subsumes this distinction; it is in a sense a unity of opposites, or of mutually exclusive attributes. The metaphysically absolute character of this distinction, or mutual exclusiveness between the attributes is proven, according to Spinoza, by the eternality and infinity of these attributes. On the basis of this, the substance in which these attributes inhere and the identity between substance and its attributes are absolute, metaphysical. Diderot transforms this metaphysical unity of substance into a very different sort of material unity. He does this by introducing a dynamic process, a real temporal development as characteristic of the one substance, the matter of the universe. The essential point of Spinoza's monism, the equal reality of the different attributes, is transformed by Diderot into the equal materiality of all that exists. The attributes of matter for Diderot, like the attributes of substance for Spinoza, have "always been in [matter, substance] together," are the modes of existence of matter, the way that matter is. But the attributes of Diderot's matter have little in common with Spinoza's attributes of substance. Matter, for Diderot,

although it is the one underlying unity of the universe, is particularistic. Spinoza's logic includes distinct, different attributes under the category of one substance. Yet this logic pertains to absolute, metaphysical attributes. Particular things or bodies are also real, also identical with substance, insofar as they are modifications or modes of substance. Yet they are finite and temporal, coming about by the endless cause and effect of Spinoza's determinism. Therefore bodies exist, insofar as extension is an attribute of substance. They have not the essence of the attribute, i.e., are not infinite and eternal. The modification, the *necessity*, or *determination*, required for the body makes it unfree, not self-caused, but determined by mechanical causation. Neither do bodies have self-motion, for Spinoza, but only mechanically imparted motion (*Ethics*, II, xiii, Schol. Axiom 2 Lemma 3). In this sense, extended material body is merely a concretization of the modifications of a metaphysical attribute. Material body is removed by several degrees from the essence of substance. It is removed further in the sense of a negation of the very essence of substance itself, i.e., it is finite, temporal, determined, in negation of the infinity, atemporality, self-causality of substance. Nevertheless, by virtue of Spinoza's monism, body is no less real or substantial, i.e., *in* substance, for all that.

Matter as an attribute of substance, pertaining to the essence of substance, is not divisible into parts. This metaphysical nature of Spinoza's matter, its 'reality' in the scholastic tradition of realism, is rejected by Diderot. The structure of matter is for Diderot more directly derived from mechanical materialism. Yet the particularist, nominalist matter of mechanical materialism lacks the metaphysically rich monism which Spinoza's philosophy possesses. It is the unity of substance with its attributes, the logic of unity of mutually exclusive attributes, that Diderot uses to overcome the atomistic mechanism, which lacked this organic sense of inner unity, whose unity was in terms of external causal interconnection only. The rich crosscurrents of Spinoza's monism, and mechanical materialism, resulted in Diderot's new materialistic monism.

DIDEROT'S SYNTHESIS OF MECHANICAL MATERIALISM AND THE DYNAMIC MONISM OF LEIBNIZ

Diderot, as an alleged deist[3], is constantly arguing in his earliest works against what is rather obviously presented as a superior antagonist. This antagonist is an atheist, whose philosophy is materialist. In the early *Pensées philosophiques* Diderot writes:

I open the notebooks of a celebrated professor and I read: "Atheists, I grant you that motion is essential to matter – what do you conclude from that? That the world results from the fortuitous course of atoms? . . ." I would avoid such reasoning with an atheist. This comparison would give him an open field. According to the laws of the analysis of chance, he would tell me, I shouldn't be at all surprised that a thing occurs when it is possible, and the difficulty of the event is compensated by the quantity of chances. [A.-T., I, 135]

This mechanist Lucretian view, which Diderot is not so subtly espousing here as a deist, deals with an atomistic universe. The basic cosmological assumption here is not in terms of substance or attribute. It is in terms of an atomic structure out of which all things are formed by the inherent motion of the atoms. This is the direct source of Diderot's materialism. In terms of Diderot's development, the rich monism which develops later is based on, and referred back to, his earlier mechanical materialism. Diderot starts as a mechanical materialist, together with D'Holbach, La Mettrie, and Helvétius. He insists on a consistent materialism, rejecting all deistic elements. The times demanded sharp, consistent formulation, and Diderot demanded this of his contemporaries. In a critique of Helvétius, he writes:

I like a clear, sharp, and frank philosophy such as is found in the *Système de la nature* and even to a greater extent in the *Bon Sens* [both by D'Holbach]. I would have said to Epicurus, 'If you don't believe in the Gods, why relegate them to the spaces between the worlds?' The author of the *Système de la nature* isn't an atheist on one page and a deist on the other. His philosophy is all of one piece (A.-T., II, 398).

But consistency alone is not for Diderot the sufficient criterion of the truth of a philosophy. He definitely does not regard philosophic or scientific formulations as 'conventional,' whose truth lies only in consistency. Commenting, for example, on the rise of organic forms out of inert matter, in his refutation of Helvétius, Diderot comments:

[This] is only a supposition which draws all its strength from the difficulties it overcomes, which in itself does not suffice in good philosophy . . . (A.-T., II, 302).

Diderot's 'proof' of materialism is not a definition, a 'supposition' overwhelmingly conclusive in its static absoluteness. It is in analyzing the very processes and development of the material world in a way which explains all phenomena materialistically, it is in the processes of matter itself, that he finds its proof; that is, he recognizes his philosophic materialism as a proof in process, not as ultimate truth in itself, the crowning point, the final truth of all philosophy. In his *Pensées sur*

l'interprétation de la nature, he writes: "Have a system – I agree to that. But don't let yourself be dominated by it" (A.-T., II, 23.)

The system of mechanical materialism was based on the divorce between physics and theology. It was not at first a universal systematic assertion of materialism. As mechanics was applied philosophically to those realms which had previously been forbidden territory, God and divine causation were further and further removed from the world. All the thisworldly universe was finally free of miracles and divinities, but like Lucretius, the deists put God 'outside' the world machine – which might have been, for them, the better part of valor. For Diderot, the advance that deism represented was significant: the physical world had come into its own. In the allegedly deistic *Pensées philosophiques,* he writes:

It is only in the works of Newton, Mueschenbroek, Hartzokher, and Nieuwentit, that satisfying proofs of the existence of a sovereignly intelligent being have been found. Thanks to the works of these great men, the world is no longer a god. It is a machine which has its ropes, its pulleys, its springs and its weights (A.-T., I, 133).

But in the same early work, the dualism of the world-machine with a God external to it is answered with: "Enlarge God: see him everywhere, wherever he is, or say that he is nowhere at all" (A.-T., I, 138). That is, the realm of physics had also to be nowhere or everywhere. A monist consistency demands, then, the basing of everything, of all existence, either in God or in the physical world.

The point at which mechanical materialism departs from the God of the 'first push,' expressed either as Prime Mover, First Cause, or Creation, is when it makes self-motion a characteristic of matter. It is also at this point that mechanical materialism begins to negate its mechanistic character. The sources and the development of this concept of the self-motion of matter, of that kind of matter which makes up Diderot's universe, do not emerge from mechanical materialism, which finds the problem of self-motion insurmountable within its context of quantitative, imparted, mechanical motion. In Descartes, motion was said to inhere in matter, but the ultimate organic identity of motion and matter was not explained in materialist or monist terms. God created matter and put motion into matter. The problem of continuously maintaining this motion in matter brought about the "retreat of occasionalism," as Abraham Edel has called it. It was a retreat in the sense that the divorce of physics and extended substance from theology

and thinking substance, the dualism of Descartes, was a step forward. It provided the necessary precondition for the development of a materialist physics, free to become universalized in a mechanistic materialism. The constant maintenance of motion in matter, which Geulincx and Malebranche proposed in their occasionalism, brought divine and spiritual causation back into the material realm of physics, and did it on the very grounds of mechanistic causation. The very limits of mechanistic materialism, of atomist causation alone, forced this resort to the intervention of God to explain motion in matter. The self-motion of matter needed something other than itself to start it.

The idealistic monadology of Leibniz served as the source of this concept of self-motion; its materialization in the hylozoism and vitalism of Maupertuis and Robinet, in eighteenth-century France, indicates the line of development of Diderot's completely materialist formulation of matter in motion.

The self-motion of Leibniz's monads was inextricably bound with the organic unity of the entire monadic system, and God was for him the all-encompassing substance. But two elements in the monadic system make it possible for it to become a more direct source of the *structure* of Diderot's materialism than Spinoza's substance. First, the pluralism of Leibniz's system: each monad was in itself a substance, an ideal self-sufficient world. At the same time, Leibniz's system was in essence not pluralistic. The internal organic interconnection of all the monads with each other, their total involvement in the whole, made this a monism which was much closer to the atomistic universe of the mechanical materialists. The particularistic character, the essential underlying nature of differentiation in the system of monads, created a more dynamic, less metaphysically static unity than that of Spinoza's substance. Secondly, the dynamism of the Monadology gave it that sense of development, process, and change, which already bespoke the negation of absolute, metaphysical categories. Force, as the underlying real existence of the monads, made the whole system a process rather than a static category in the abstract. This too, in terms of mechanical materialism, was closer to the world of coursing atoms than Spinozism.

Diderot had a keen understanding of this concept of process. His materialist monism, though grounded in the unity of matter with its attributes in terms of the logic of Spinoza's monism, did not work within the categories of scholastic theology, as Spinoza's did.[4] Rather, Diderot's monism dealt with the categories of that physics which had been divorced

from theology in Cartesian dualism. Body and motion were the first essential categories of mechanical materialism, whereas they were unessential, finite modifications of the essential categories of substance and attribute for Spinoza. In Leibniz, body and motion – monad and force, in his idealized terms – are the essential categories. They were stripped of their idealist character and translated into materialist terms by virtue of their close formal and structural similarity to the categories of mechanical materialism. The dynamic, self-sufficient mode of existence of the monad is made the mode of existence of the material atoms.

In Leibniz unity is conceived in terms of force and process. But it is the force and process of particulars that make up this unity. Force, process, and self-motion do not exist as substantive categories apart from their existence in particulars. This identity of the mode of existence of these particulars, that is, their self-motion etc., with the particulars themselves, is a key element of Leibniz's monism that is present in Diderot's materialistic monism. For it is on this particularistic unity that is based the rejection of the category of substance as a support, a ground, an abstract and undifferentiated thing-in-itself. Substance as real, undifferentiated essence, as God, is reduced or rather transformed into the real heterogeneous existence of particulars, outside of which nothing else exists. Spinoza, by basing God immanently in all, pointed to this disintegration of the concept of a substance-in-itself, prior to, or separate from, its particular attributes. By rejecting the substantive God in Leibniz and accepting only the self-sufficient monads, Diderot was thus able to find a way of overcoming the mechanists' lack of an organic monism.

THE MOLECULE: FROM MONAD TO MATTER

The characteristic of mechanical materialism historically is that it starts with a building-block and constructs the universe out of this material element. Whether with Epicurus and Lucretius, or with Gassendi, Bacon, Hobbes, Newton, Locke, or the French materialists of the eighteenth century, a material atom or molecule is the basic structural element of the universe. Particulars exist, of which the unity of the world and its forms are constituted. Idealism starts from the opposite pole, with an abstract metaphysical sense of this unity as God, Spirit, Substance, Idea, which gives rise to its particular expressions, its things and bodies; these

emanate from this abstract ideal unity through its very nature. It, rather than the particulars, is the pre-eminently real existence. Hegel, from the point of view of his idealist dialectic, characterizes this difference between the mechanical materialists and the idealists, when he says, speaking of French eighteenth-century materialism:

Here the Idea in its infinitude is not itself the object of knowledge, but a determinate content is raised into the universal (*History of Philosophy*, Part 3, Introd.).

For Diderot, the monism of the idealist Leibniz is an inversion of material reality. The qualities of self-motion and force of the ideal monad Diderot recognizes as the qualities of the material molecule, of the elemental matter of the universe. Is this a metaphysicisation of matter, attributing to it those ideal qualities which are posited only metaphysically? Only if one assumes the equal validity, or non-validity, of idealism and materialism. From Diderot's point of view as a materialist, the monad is not an ideal fiction merely, but an idealistically inverted truth about the material world. How can a consistent materialism develop on the basis of a universe made up only of material particulars? Only if all natural phenomena can be explained in terms of these particulars themselves. If there arises the need, on the part of mechanical materialism, to call in a deist God or a non-material First Cause as an explanation, that is a sign that the reach of material explanation is insufficient; that the elemental particulars of the material universe have not yet been fully understood, their properties not fully realized or discovered.

Maupertuis, a contemporary of Diderot, takes the first steps in the materialization of the Leibnizian monad. In his letter on the monads, he states: "The monads, according to their principles, cannot be anything but the primary elements of matter, endowed with perception and force."[5] He is ready to endow these elements of matter with all the properties necessary to explain natural phenomena. But in this way, he rejects the possibility that these properties may in some cases arise and develop only at certain points in matter's own development into different qualitative levels. Perception and intelligence become for Maupertuis essential properties of matter. In his *Système de la nature*, he writes:

The more phenomena there are to explain, the more it is necessary to charge matter with properties . . . but if, with all the properties that have been admitted here, it is not possible to explain the formation of organized bodies, it would become necessary to admit of more properties, or rather, to recognize the properties that are there (*Oeuvres*, II, *Système*, xxv–xxvi).

The materialist side of Maupertuis is in the last phrase. The properties
are not merely instruments of explanation, conventions; they actually
exist in matter waiting to be discovered, under the pressure of new
scientific discovery, new phenomena to be explained. Maupertuis regards
the material molecules as endowed, not as a principle of explanation, but
as a rule of their existence itself, with the least possible properties needed
for their activity. This he derives from Leibniz's 'least possible force
necessary' principle. He writes: "In the explanation of these phenomena,
we have only one rule to observe: that is, we will use the fewest possible
principles and the simplest principles possible" (*Oeuvres* II, *Système*,
xxiv). He takes, therefore, the two substances of Descartes, and with
Spinoza, asks: "Though thought and extension are distinct properties
can we not find them existing in the same subject?" (*Oeuvres*, II,
Système, xxiii).

Maupertuis is ready to "accord a principle of intelligence to matter,"
in his rebuttal of Diderot's criticism in *Pensées sur l'interprétation de la
nature*. He states, in the *Système de la nature*, that when God created the
universe:

. . . He endowed each of the smallest parts of it [matter], each element with some property
similar to what we call in ourselves desire, aversion, memory; the formation of the first
individuals being miraculous, those which succeeded them are no more than effects of
these properties (*Oeuvres*, II, *Système*, xxxi).

Ernst Cassirer criticizes this in his *Philosophie der Aufklärung* as a
vergröberung, a vulgarization of Leibniz's idealism:

Leibniz's spiritualism is vulgarized here to a vague and unclear hylozoism. Matter as
such is ensouled; it is endowed with sensation and desire, with decided sympathies and
antipathies. And such an instinct is given to every part of matter.[6]

What Cassirer criticizes as a vulgarization of Leibniz's idealism is, in a
sense, the inconsistency of Maupertuis' materialism. The monad has
become a material element. But the principle of divine endowment of this
element with properties and Maupertuis' inability to break through the
Cartesian dualism of physical and psychical realms in a monist way lead
him to the idealist element in his philosophy:

One might admit in matter properties of another order than those which we call physical.
. . . One might accord matter some degree of intelligence, desire, aversion, memory. I
believe I see the necessity for this. The formation of any organized body will never be
explained by the physical properties of matter alone (*Oeuvres*, II, *Système*, xxviii).

In spite of this idealist element, Maupertuis has started on the road towards the materialization of Leibniz's dynamic monism. Luppol, in his book on Diderot, notes this:

Maupertuis . . . took the first step in transforming the incorporeal monad into a physical molecule, but in according it a superior psychic activity, he separated himself from nature, or, if you please, deviated towards idealism.[7]

Diderot in his Encyclopedia article *Hylozoism*, also sees in the sensitive soul attributed to each particle of matter a sort of divinity, ". . . a blind goddess whom they [the hylozoists] call nature." Robinet, another contemporary of Diderot, develops a theory in which the forceful potential monad is transformed into a 'germ,' whose unfolding produces its being, actualizes it, works itself out as the form which we perceive. For Robinet there is a deistic Prime Mover, but one closer to the Leibnizian than to the mechanistic concept of a first force. After the first and permanent *impression* made upon the world by God at the Creation, ". . . the universe goes on living, moving, and perpetuating itself . . . the whole maintains itself through mutual correspondence of its parts."[8]

The sense of unity arising out of particularity in this pre-established harmony that has its roots in Leibniz, is here on the way to achieving its completely particularistic expression. "The whole is maintained by the mutual correspondence of its parts," says Robinet. The unity is not *a priori* substantive existence or reality, but is rather a constitutive unity of particulars. The particulars in this case have an inner organic development, in an Anaxagorian sort of atomism. Dog-germs develop into dogs, man-germs unfold into men. The reality of these particulars is a sort of physiological potentiality which exists in this particularistic, even pluralistic manner – pluralistic in the sense that there is no chain of being, no continuum as in Diderot, but an absolute and eternal nature in each 'germ'. Maupertuis too exhibits something of this vitalist emergence in the theory that all forms spring from variation of their prototype, their 'animalcule.' Thus Maupertuis speaks of ". . . each animalcule which will form a man, or which is, rather, the man fully formed" (*Oeuvres*, II, *Système*, lvii). According to this emergence in Robinet and Maupertuis, no one kind of monad, germ, animalcule, can give rise to any other kind. The germs are self-subsistent and disparate. It remains for Diderot to posit the actual transition and transformation of species, in distinction to these metaphysically eternal germ-substances.

DIDEROT'S MOLECULE

These philosophic forbears and contemporaries of Diderot developed either a monism which was not materialist, or a mechanical materialism which was not monistic. The active dynamic side of matter, which dealt with organic inner change and development, inherent force and motion, had been developed not by materialism, but by idealism. The ontological materialist basis of Diderot's monism did not have its source in the metaphysical tradition; its source was mechanical materialism rooted in the materialist science of the time.

The most striking exponent of a consistent materialism in Diderot's time was Diderot's close friend, Baron d'Holbach. His was a most forthright exposition of a total and uncompromising materialism. The universe for him was nought but matter in motion. All phenomena, all that exists, is made up of matter in motion; is formed, maintained, destroyed, as matter in motion. The underlying monism here was a simple enough one: the world was all material. Therefore it was monistic in its materiality. This simple, almost too simple materialism, carries within it the basic premise of Diderot's materialist monism; not a formal statement of unity, but a thoroughly ontological, materialist one. D'Holbach's materialism, however, is limited by the mechanistic logic of quantitative combination and interconnection alone.

Diderot's matter has motion as an inherent property. It is not endowed with motion; it is not a ground in which motion is put. Matter itself is uncreated, eternal, its motion is its essential mode of existence. This motion is not Spinoza's entirely mechanistically imparted motion given to a body by an external cause in the endless chain of cause and effect. This self-motion of matter is rather a transformation by Diderot of the self-motion of the Leibnizian monad. He adopts the monad at first on formal grounds, not as either material or ideal, rather assuming for it a materiality without seeing the full significance of this assumption, the full meaning of the basic transformation he has made ontologically. In his article on Leibniz, he describes the monads as

. . . themselves the cause of their internal actions. These are like incorporeal automatons. What difference is there between these beings and Hobbes' sensitive molecule? I don't see any (A.-T., XV, 457).

In equating the incorporeal self-moving monad with Hobbes' 'sensitive molecule,' the world of difference between the incorporeality of the one

and the corporeality of the other is not noted. He assumes for objective idealism an uncritical formal identity with materialism. This evidence in some of his early work has led Isidore Sherman, for one, to divorce Diderot's monism from his materialism, and has led Janet, by the same error, to label Diderot a Leibnizian. Diderot is much more aware later on of the essential difference. It is enough for him at this point, however, that the monad is real. He comments on the internal action of the monad:

. . . the changes in the natural monad arise from an internal principle because no external cause can exert an influence upon it. *In general there is no force whatever which is not a principle of change* [My italics.] (A.-T., XV, 456).

He sees then in Leibniz the interrelation between the principle of internal action, inherent force and motion, and the phenomena of change. Moreover, this change takes place not as an external rearrangement of parts, as a mechanical change, but in the absence of parts. This absence of parts, excluding mechanical change and making change itself a function of the monad, leads Diderot, in materializing this monad, to speak of the inherent properties in the hylozoist terms of Maupertuis, which are derived from Leibniz: appetition, desire, perception, etc. Diderot writes, in the article on Leibniz:

All natural change takes place by degrees. One thing changes while another remains unchanged. Therefore in substance, there is a plurality of affections, of qualities, and relations, though there is an absence of parts. . . . The action of an internal principle causes mutation, or the passage from one perception to another, that which is called appetite . . . (A.-T., XV, 456).

The point on which Diderot differs sharply from the hylozoist sense of appetition – ensouled molecules – is Maupertuis' positing of the mutual exclusion, in an absolute sense, of the molecules. Maupertuis still maintains the 'hard body' of mechanical materialism, and his ensouled molecules have only external interconnections, despite their internal appetitions. Each molecule, then, is organically an isolate, much like the Democritean atom. The pre-established harmony of Leibnizian substance is lost in the absolute isolatedness of Maupertuis' material monad. Diderot's understanding of the universe as an organic unity made up of particular elements is misstated by Maupertuis, whose mechanist limits keep him from seeing any continuity or unity in a discontinuous atomistic universe. He writes in answer to Diderot's critique:

The parts of matter being always distinct, and the one never being able to become the other, however close they may be, however intimately they are brought together, they will never make of the universe more than an apparent continuum. The difference between this continuum and the discreteness of the parts of the universe consists only in the greater or lesser distance between these parts, is only the effect of this distance between parts on our senses. . . . Microscopes have been able, or will be able, to make perceptible to us the distances between the parts of the most compact bodies; our sight and our touch may be mistaken, but for our reason, no material is continuous. [*Oeuvres*, II, 204]

For Maupertuis, the continuity or unity of the material universe is understood only in terms of mechanist unity; that is, the unity of the indivisible atom, of a solid block. This kind of unity of the material universe as a whole is therefore rejected by him. The metaphysical unity of substance, or of the monads, is for him, after all, metaphysical, and Diderot is proposing not an ideal, spiritual unity, but a material unity of all being. Maupertuis fails to grasp the sense in Diderot of a unity of discrete particulars – in this sense, a unity of opposites. For Diderot, one molecule does not become another molecule, but qualitatively the totality · of molecules becomes a whole at a level no longer mechanistic. It is Diderot's concept of the development of qualitative levels by the constant change in the universal matter in motion that permits him to break through the limits of mechanism within which Maupertuis and all the mechanists are bound.

In the *Pensées sur l'interprétation de la nature*, Diderot expounds the theory of Dr. Baumann of Erlangen (Maupertuis' pseudonym) at length. He comments on the "terrible consequences of this hypothesis." Diderot asks Maupertuis:

. . . if the universe, or the general collection of all the sensitive and thinking molecules forms a whole or not? If he [Maupertuis] answers that it does not constitute a whole at all, he is destroying with one word the existence of God, by introducing disorder into Nature; and he destroys the basis for philosophy in breaking the chain which links all beings. If he agrees that the universe *is* a whole . . . then . . . as a result of this universal copulation, he will have to admit that the world is similar to a great animal, to a soul; that the world thus capable of being infinite, the soul of this world can be – I'm not saying that it is, mind you – an infinite system of perceptions, and the world can be God. [A.-T., II, 48]

In materialist physics, Diderot had pointed out, the world is no longer a God. Yet this later materialism is given its monistic meaning in this 'consequence' that Diderot derives from Maupertuis: the world is God. But this traditional language is used by Diderot to draw the conclusion of

a complete atheism, a complete immanentism, in which the world is based in God, or rather, God is based in the world – the very atheism that Spinoza was accused of. Maupertuis rebukes Diderot for drawing this conclusion from his *Système de la Nature:*

If one were less convinced of the piety of the author of the *Pensées sur l'interprétation de la nature*, one could be led to suspect that his object was not to destroy [my] hypothesis, but to draw from it these consequences which he labels 'terrible' [*Oeuvres*, II, 197].

Maupertuis, in combining the elements of the monadology with mechanical materialism, lays the ground nevertheless for the 'terrible consequences' of a monist development of materialism. The fuller, richer monism of Diderot completes this development by negating the isolatedness, the limited external interconnection, which characterizes mechanism. This monism remains materialistic by making the basis for unity not an ideal substance, or a God, but matter itself. Berkeley, with a sharp idealist sense of the 'danger' of this consequence, of this kind of monism inherent in materialism, says, in his *Commonplace Book:* "Matter once allowed, I defy any man to prove that God is not matter."

DIDEROT'S PHILOSOPHIC PRINCIPLES OF MATTER AND MOTION

Diderot develops his analysis of motion in the sketchy but brilliant *Principes philosophiques sur la matière et le mouvement*. Diderot's analysis of motion, the basis for his dynamic monism, of his universe in flux, is based on the attribution of inherent force and motion to bodies, in the Leibnizian tradition, and on a relationism which precludes any absolute rest, in the Cartesian tradition. In his *Principes philosophiques* he argues against Maupertuis' analysis of motion in the latter's *Essai sur la cosmologie*. Maupertuis divorces the Leibnizian force inherent in the monad from motion. He states that one must not confuse force with motion or regard force as the cause of motion. Force, he says, has become an obscure term. The only measurement of force is by its apparent effects [*Oeuvres*, I, 30ff]. Motion is not an essential quality of matter, but only an apparent effect of force. Therefore absolute rest exists as the absence of this effect. Motion may or may not be. Matter is indifferent to motion or rest.

We see some parts of matter in motion; we see others at rest. Motion is therefore not an essential property of matter; it is a condition in which matter either may or may not

find itself, and which we cannot see deriving from matter itself. The parts of matter which are in motion have therefore received it from some external cause, which at this point is unknown to me, and since [the parts of matter] . . . are in themselves indifferent to motion or to rest, those which are in motion will remain in motion until something changes their condition. [*Oeuvres*, I, 32]

This fully mechanistic analysis of motion makes it a conditional state of matter. Diderot makes motion an essential property of matter, denying absolute rest:

"I do not know in what sense philosophers have supposed that matter is indifferent to motion and to rest. . . . Everything is in a relative state of rest in a boat being tossed about by a tempest. Nothing there is in a state of absolute rest however, not even the molecules constituting the boat and the bodies contained in it (A.-T., II, 64–65).

Diderot answers Maupertuis, then, that the appearance of absolute rest is only appearance; that the essence of the material world is motion, and that relative motions give the appearance of rest in some parts of matter and of motion in others. But it is not meant here that motion is essentially relational: its appearances only are relational. Diderot does not posit a formal, quantitative relationalism, but bases his relationalism on a qualitative property of matter. Moreover, for him, motion or action are not caused by force, but force and motion are conceived of as a unity. If force is inherent in matter, then it is continuously active; its action is motion. This unity of force and motion is similar formally to the unity of substance and attribute in Spinoza. The attributes in Spinoza's substance are essential, eternal, uncaused by substance, not in any way created. Substituting force for substance, Diderot says that motion has this same essential identity with force.

"Body", according to some philosophers, "is in itself without action and without force." This is a terrible falsity, contrary to all good physics and to all good chemistry. In itself, by the nature of its essential qualities, whether body is considered as molecule or as mass, it is full of action and force. "In order to represent motion to yourself," they add, "it is necessary to imagine a force acting on matter." This isn't at all the case. The molecule endowed with a quality proper to its nature is *in itself* an active force. It exerts its force upon another molecule, which exerts *its* force in turn upon it. All the paralogisms I noted above are related to the false supposition of homogeneous matter. You who can imagine matter at rest so well, can you imagine fire at rest? Everything in Nature has its own diverse action, like this mass of molecules which you call *fire*. In this mass that you call fire, each molecule has its own nature, its own action . . . [A.-T., II, 65–66]. If they [these philosophers] do not conceive any more of a tendency to rest than to motion, it is apparently because they regard matter as homogeneous; it is because they reason about the relative rest of one aggregate in relation to another aggregate. It is because they forget that while they are reasoning about the indifference of bodies to motion or to rest, the block of marble is proceeding to its dissolution. It is

because they negate by thought the general motion which animates all bodies, and the particular actions of one upon the other which destroys them; it is because this indifference, as false as it is in itself, and only transitory, will not prove the laws of motion to be erroneous. [A.-T., II, 65]

The Heraclitean comparison of matter in flux to fire points out the dynamic particularism of Diderot's matter. Matter is spoken of, not as itself a substantive block-like entity, or an atom-like solid continuum, which Maupertuis imputes to him. Rather it is itself the totality of the particular molecules. The motion of the particulars is self-motion; yet this does not preclude cause and effect. Diderot's pre-established harmony, so to speak, is not teleological like Leibniz's. This harmony, moreover, is nothing but the totality of the motions of the particular molecules. Here too, Diderot's monism starts from the particulars rather than from the abstract, teleological Absolute of idealist monism.

Diderot denies the homogeneity of matter, of the material molecules. That is, he not only asserts the particularity of the structure of the material world, quantitatively, but also its heterogeneity, qualitatively. This heterogeneity is essential, i.e., it concerns the inherent force and action of the molecules. Thus motion itself is not a homogeneous phenomena; it is qualitatively differentiated, which makes possible the appearance of relative rest and motion. What of the interconnection between the molecules? Basic to mechanical materialism is the external interconnection of all the atoms. There is an endless chain of cause and effect, an endless imparting of an original amount of motion, resulting in translation, change of place. This external interconnection exists, says Diderot. But interconnection for him is not merely the imparting of local motion to perfectly hard bodies, not merely translated motion according to the laws of chance, or the chance course of Lucretian atoms swerving through the void. The weakness of this concept of mechanistic motion is that at some point, at a 'beginning,' this motion was put into inert matter, from some source outside matter. This matter can never be identified with motion, is never more than a ground, a receptacle, which merely holds this packet of motion it has received, until it bumps into another particle of matter and passes off the motion to it. Therefore, there is either a Prime Mover, and some eternally predicated laws of motion external to matter to which it somehow conforms, or there is a continual occasioning of motion from some outside source, or motion and matter are two separate substances, and the material world is a combination of matter *and* motion, with some divine means needed to

connect the two substances. For Diderot, however, the unity of matter
and motion is an eternal, uncreated unity: motion is merely the mode of
the existence of matter. When two such inherently motive molecules do
'bump' into each other, a translation of force takes place completely and
begins and ends there with a change of place. But this translation is the
effect of self-motion. Therefore, Diderot speaks of two sorts of *Nisus:*

> For a molecule to be moved, a force, an action is necessary, they tell us. Yes, either
> external or internal, inherent, essential, intimate to the molecule, constituting its nature
> as a fiery, watery, nitrous, alkaline, or sulphurous molecule. Whatever this nature may
> be, force flows from it, action of the molecule outside of itself, action of other molecules
> upon it. The force which acts upon the molecule becomes exhausted. The internal force
> of the molecule never becomes exhausted, it is immutable, eternal. These two forces
> can produce two sorts of *Nisus:* the *first,* a *Nisus* which ceases; the *second* a *Nisus* which
> never ceases. Therefore it is absurd to say that matter has a real resistance to motion
> [A.-T., II, 66]

Imparted or violent motion, then, comes to an end. It is only a form of
the eternal self-motion of matter. The interconnections of these
self-motions are what make up the laws of motion of the universe. They
are interconnections of the internal properties or qualities of matter.
They are *qualitative interconnections.* This is the meaning of Diderot's
'heterogeneity' of matter. It is the characteristic of the different motion
of each molecule that makes one 'sulphurous,' the other 'nitrous.' Motion
is not merely quantitative here, but the quantitative and qualitative
aspects of motion are interrelated. Therefore the laws of motion are not
merely quantitative, but qualitative as well; the laws of motion therefore
inhere in self-motive matter. On this basis Diderot criticizes the merely
quantitative proportionality between force and matter, energy and mass:

> Yes, without a doubt, if one compares a homogeneous aggregate to another aggregate
> of the same homogeneous matter, then, when one speaks of the action and reaction of
> these two aggregates, their relative energies are in direct proportion to their masses.
> But when it is a question of heterogeneous aggregates, heterogeneous molecules, these
> same laws do not apply. There are as many diverse laws as there are varieties of the
> inherent, self-contained force of each elementary molecule constitutive of bodies.
> [A.-T., II, 67]

Though these laws vary with each elementary molecule, Diderot does not
preclude general laws of motion which subsume several or all of these
diverse laws. As we have seen, he foresees, with the development of
experimental physics, such a generalization. His attack on the concept of
homogeneous matter and homogeneous force is, in fact, his attack on

mechanism. Diderot therefore establishes in opposition to this a qualitative molecular particularism. He does not state at this point the developmental dynamic view of the interrelation of quantity and quality, which he states in his physiological writings, but the logic of this later evolutionism is implicit here. He sees the possibility of qualitative change in an aggregate and criticizes the logic of mechanical physics:

> What are these philosophers doing, whose errors and paralogisms I am refuting here? They are concerning themselves with one single and unique force, which is perhaps common to all the molecules of matter. I say *perhaps*, for it wouldn't surprise me at all if there were in nature certain molecules which, when joined with others, would make their resulting combination lighter. Every day in the laboratory, we make one inert body turn another inert body into a gas; and those who, considering all action in the universe to be only that of gravity, thence conclude the indifference of matter to rest or motion, or rather the tendency of matter to rest, believe that they have solved the question, whereas they have not even touched its fringes. [A.-T., II, 69]

Diderot speaks of three kinds of earthly action or force on the 'globe we inhabit'; gravity, the internal force of each molecule, and the action of each molecule on every other molecule. The result is a vector characterizing the motion of each molecule. Essentially, though these are the three kinds of forms of force, all force is ultimately inherent, all motion is self-motion. All effects follow from this self-motion. Diderot leaves no room for other sources of motion than motive matter itself. The total interconnectedness which this inherent force or motion implies is stated by Diderot thus: "An atom moves the world. Nothing could be truer. It is as true as an atom moved by the world; since the atom has its own force, it cannot be without effect" (A.-T., II, 67).

FROM INORGANIC TO ORGANIC MATTER: MAUPERTUIS AND LA METTRIE

Diderot speaks of the qualitatively different laws of motion and force of different aggregates of molecules. The concept of aggregate plays an important role in Diderot. The aggregate is the temporal, conditional organization of molecules, having certain qualitative characteristics of its own. The totality of the universe is a totality of aggregates. Forms are aggregates of molecules.

The problem of the origin of forms was a key problem for the mechanical materialism of the time. Forms are accounted for only by the different combinations of material atoms. For D'Holbach, for example,

the mind is merely a certain organization of atoms. It is on the basis of this purely quantitative atomistic mechanism that mechanistic material-ism has been traditionally rejected, on the grounds that this atomism alone could not account for such refined qualities as sensibility, intelligence, mind. Cartesian dualism rests on this. Thinking being is seen as qualitatively distinct from extended being, and it is this qualitative gap that a quantitatively-limited mechanism, the mechanism of homo-geneous matter, cannot bridge. For Descartes, the animal is a machine, since all activities of organism as such can be accounted for mechanistically. The qualitative difference between inorganic and organic matter is ignored in favor of the greater qualitative difference between thinking substance and extended substance. The animal has no soul, and past this, according to Descartes, animation, intelligence even, is possible mechanistically.

Let us examine here some of the roots of Diderot's transformism, so-called, his materialistic theory of the development of forms and qualities in matter, which provides a bridge for this gap. La Mettrie, being more consistent in the mechanist tradition than Descartes, states in *L'Homme machine* that all human activities and qualities can be accounted for mechanistically.

The human being is a machine which winds its own springs. It is the living image of perpetual motion. Nourishment keeps up the movements which fever excites. . . . Everything depends on how the machine is running.

The qualitative differences are here obliterated altogether for the sake of a materialistic consistency within the limits of mechanism. Yet even for La Mettrie, the question of organic life demands more than mechanism alone for its own solution. He turns to a theory similar to the 'germs' of Robinet and the 'animalcules' of Maupertuis. In a very interesting and neglected work, the *Système d'Epicure*, La Mettrie speaks of *semences*, seedlike essences in the air, which come to fruition under certain material conditions as this or that form of organic life. Both plants and animals, i.e., certain conditional forms of the mechanical organization of matter, abstract their appropriate *semences* from the air. Thus, these *semences* correspond to a sort of germ-plasm, which seeks its fruition in the proper ovaries and testes. Man developed thus from the earth at a certain appropriate time, under certain material conditions. The earth served, says La Mettrie, as a uterus for man, attracting to it the man-seeds. Why isn't the earth still producing men? asks La Mettrie. Because it is too old to be a mother any longer, he answers.

In essence, La Mettrie is calling in a qualitative element, the *semences*, to account for qualitative differences. His matter is heterogeneous in the sense that certain organizations of matter, by acquiring these *semences*, are qualitatively distinct. But the principle of qualitative distinction does not, at the level of organic or living matter, inhere in the elementary particles of matter themselves. It pertains only to certain combinations of matter. Qualitative difference, then, depends on quantitative organization, plus a qualitative element external to this combination. It does not depend on organization alone. La Mettrie, the ultra-mechanist, finds, at least in the *Système d'Epicure*, a need for the qualitative *semences* to explain the origin of species. He is not a vitalist in the same sense as Maupertuis. For him matter as such is not permanently endowed with life, or perception, or intelligence. As we have seen, Maupertuis was ready to endow matter, even each particle of matter, with all these properties. Therefore these properties or qualities become essential eternal qualities of matter. Maupertuis calls them "non-physical," since he cannot see the possibility of "explaining the formation of any organized body by the physical properties of matter alone" (*Oeuvres*, II, *Système*, xxviii).

By physical properties are here understood force or action, length, mass, etc. The qualities added are, then, immaterial, psychic qualities. Yet for Maupertuis, the mechanism whereby these qualities in the molecules achieve this or that organic or inorganic form is by combination. This is a combination of *more* and *less* active parts of matter. His homogeneous molecules differ, not in quality, but in degree or intensity. For example, this molecule has more perception, is more active than another. "The less active parts of matter formed the metals and the marbles; the more active parts formed the animals and man" (*Oeuvres*, II, *Système*, xlix).

There can be no growth or diminution in the totality of these qualities, i.e., they follow a sort of law of the conservation of energy, in terms of the conservation of perception: human qualities take on the aspect of eternal, essential qualities in general:

Perception being an essential quality of elements, it doesn't appear that it can either perish, diminish, or grow. Indeed, it can receive different modifications by the different combinations of elements; but it must always form the same total in the universe, though we cannot trace it or recognize it.

We cannot trace it, says Maupertuis, because

... each element in its union with the others, having confounded its perception with
theirs and lost the particular sentiment of *self*, the memory of the primitive state of the
elements is not available to us, and our origin must remain entirely lost to us" *(Oeuvres,*
II, *Système*, liii).

This quantitative combination of perceptive molecules results, not in
essentially, qualitatively distinct forms, but in forms distinguished by the
intensity of degree of their common qualities. The reason for certain
combinations, or for combinations resulting in certain forms, is entirely
teleological in Maupertuis, taken over directly from the idealist
pre-established harmony of Leibniz. This is described by Maupertuis as
"that view in which the elements themselves, endowed with intelligences,
arrange and unite themselves to fulfill the views of the Creator"
(Oeuvres, II, *Système*, lxviii). La Mettrie on the other hand, rejects all
teleology or final cause:

All reasoning . . . based on final causes is frivolous.[9]

The elements of matter, by reason of the continual agitation, and their intermixing,
have brought about the formation of eyes. It was just as impossible not to see with them,
as it is impossible not to see oneself in a mirror. . . . The eye finds itself the mirror of
objects. Nature no more meant to make the eye in order for it to see, than the water of a
pool means to serve as a mirror to the simple shepherdess. *[Oeuvres phil.* I, xviii]

Nature makes silk the way the Bourgeois Gentilhomme speaks prose, without
knowing it; it is as blind when it gives life, as it is innocent when it destroys it. *[Oeuvres
phil.* I, v]

Having made, without seeing, eyes that see, it [Nature] has made, without thinking,
a machine that thinks. *[Oeuvres phil.* I, xxvii]

Having rejected teleological determination of forms, La Mettrie must
develop a theory of the origin of forms as the result of the organization
of matter. But he does not achieve a complete unity between such a
mechanist, quantitative combination and the rise of qualitatively distinct
forms. Therefore, his monism is not consistent, and the *semences* have to
be called in. We have, then, in Maupertuis and La Mettrie, two
approaches to the origin of forms which bear on Diderot's treatment of
this problem: one, Maupertuis', states that the combination of molecules
common in their qualities, physical and psychical, differing in the
quantitative intensity of these qualities, and by a vectorial addition or
mixing of these qualities, results in a dominance of one or another of
them, in accordance with the views or teleological 'end' in the mind
of the creator; in the other, La Mettrie's approach, the combination of
molecules merely quantitatively establishes the material conditions for

the acquisition of the qualitative *semences* in the air; the *semences* work out their inherent nature, their germinal characteristics, through matter. Diderot has the closest affinity to La Mettrie here. For La Mettrie posits a certain organization of matter as the condition for the rise of qualitative forms. Yet La Mettrie cannot relate the quantitative combination with qualitative change and cannot see the transformation, monistically, of quantity into quality. It is this 'transformation' which is at the roots of Diderot's transformism, his theory of the evolutionary development of forms, and the transformation of inorganic into organic matter.

FROM INORGANIC TO ORGANIC MATTER: CHANGE AND DEVELOPMENT: EVOLUTION AND HEREDITY

Diderot has been called a mediator between the vitalists and the mechanists by Sherman, in his study on *Diderot and Philosophic Naturalism*. To 'mediate' between two mutually exclusive views means at best eclectically to combine the two. Diderot does not mediate, but transforms both of these positions radically. He grapples with these two views, as two possible explanations of the origin of qualities, of sensibility in matter, but he rejects any eclectic attempt to combine them. He criticizes Helvétius for this:

If, starting with the single phenomenon of the sensibility of matter, either as a general property of matter, or as a result of its organization, Helvétius had derived clearly the operations of the understanding, he would have done a new, difficult, and good thing. I would more greatly esteem one who, empirically or by observation, could rigorously demonstrate either that physical sensibility was as essential a quality of matter as impenetrability, or who deduced it unconditionally from the organization of matter.

I urge all physicists and chemists to investigate what *animal*, *sensible*, and *living* substance is. In the development of the egg, and in several other operations of nature, I believe I see how an apparently inert but organized matter passes by purely physical agents from the state of inertia to the state of sensibility, but the necessary liaison of this passage escapes me. [A.-T., II, 301]

Diderot states the two theories, but he puts the burden of proof upon the physicists and chemists, upon scientific analysis, and not upon speculators. He clearly states the difficulties of each theory:

It is clear that the notions of matter, of organization, of motion, of heat, of living flesh, of sensibility, and of life, are still very incomplete. One must agree that the organization or coordination of inert parts does not at all lead to *sensibility*, and the theory of the

general sensibility of the molecules of matter is only a supposition which draws all
its strength from the difficulties it overcomes, which in itself does not suffice in good
philosophy [A.-T., II, 302].

In the case of mechanism, Diderot, with the more astute mechanists, as
well as with the opponents of mechanism and the vitalists, sees
mechanical organization itself as insufficient to explain qualitative
change. In the case of vitalism, the rise of organic and sensible forms in
matter is based upon a metaphysical assumption which is pure
speculation, unsupported by any scientifically investigable hypotheses.
The rejection of mechanism is based, justifiably, upon a realization of
the qualitative distinction between the merely 'physical,' the 'sensible,'
and the 'intelligent.' The superficial assertion of an identity, merely
abolishing this distinction, or overlooking it, is 'bad philosophy.' This
sense of the distinction, the opposition, the qualitative differences
between the physical, the sensible, the intelligent, Diderot applies in
criticism of the identity which Helvétius states in the phrase, "Sentir,
c'est juger." Although Helvétius' context is the realm of already sensible
matter, where the differentiation is between sensation and judgment,
Diderot's sense of qualitative differentiation applies here as well as to the
differentiation between organic and inorganic matter:

Isn't the distinction between the physical and the moral as real and solid as that between
a feeling animal and a thinking animal. . . . Wherefore it is important not to make of
thinking and judging two perfectly identical operations (A.-T., II, 103).

Seeing the real distinction between the organic and the inorganic, the
physical and the mental, is characteristic of Diderot's method, of his
grasp of the phenomenon of opposition, or what Cassirer calls his
understanding of the 'antinomies.' His monism subsumes this
opposition. Diderot seeks out the 'necessary liaison' in terms of matter
itself, without the idealism present in vitalism. He finds his explanation
based on the continual motion, the 'fermentation' of the universe of
matter. In Leibniz, he finds that the principle of motion is in itself the
principle of change. Motion is not instantaneous. It is continuous and
successive; it is temporal. The change proceeding from this motion
therefore has *time* as its essential condition. The particular, different
motion of heterogeneous matter gives rise to all these transformations
and forms of nature, in a temporal development.

I see everything in action and reaction; everything destroys itself in one form; every-
thing recomposes itself in another; sublimations, dissolutions, and combinations of all

species; phenomena incompatible with the homogeneity of matter; from which I conclude that it is heterogeneous, that there is an infinity of elements in nature; that each of the elements by their diversity has its particular, innate, eternal, indestructible, unchanging force; and that these forces inherent in bodies have their effect outside of these bodies, which gives rise to the general motion, or rather, the general fermentation of the universe. [A.-T., II, 68–69]

Thus, this universal motion gives rise to forms from its own nature, not as the result of any predestination, or teleology lying outside it. But how do the forms arise and pass away? The 'fermentation' of the universe is a principle of change, like that of the chemical transformations that take place in the process of fermentation. The monism of all the forms rising out of the combination and organization of matter is based on the material unity of qualitatively different aggregates of molecules. That is, the forms are distinct, yet at the same time, they pass into each other. The agent of this passage is not outside the physical properties of matter itself. Here is the basic difference between Diderot and the vitalists: although Diderot seems to equally posit the sensibility of matter as either a general property of matter or as a product of its organization, he sees that the mechanist limits of either proposition require a metaphysical attribution of the quality of sensibility, or a reduction of sensibility to the level of the mechanical motion of molecules at a single quantitative level. In other words, the very limits of mechanism itself give rise to vitalism, and subsequently, to idealism. Diderot's understanding of the limits of mechanism and how they give rise to vitalist idealism has been variously misinterpreted. He has been called, therefore, a 'mediator' by Sherman, a 'vitalist' by Venturi, and an 'idealist' by Caro. And further, he has been meaninglessly characterized by Lange as an "idealist whose materialism carried him away." The key to this misinterpretation is a passage in the *Entretien entre d'Alembert et Diderot:*

D'Alembert: . . . other obscurities result from the rejection of [the hypothesis of God] because if this sensibility which you are substituting for Him is in general, an essential quality of matter, then it necessarily follows that a stone senses.
Diderot: Why not? [A.-T., II, 105]

Taken out of context, this certainly would make Diderot as clearly a hylozoist and a vitalist as Maupertuis. But to take this passage out of context completely twists its meaning, for its true sense becomes very clear in the paragraphs that follow. Does a stone, as a stone, feel or sense? No, the stone, through the motion of its constituent parts, may go

through changes which will transform it into something that can feel.
Motion here in the stone and its concomitant change are the means by
which this potentiality of sensibility in the stone becomes actualized
sensibility. Diderot gives a well-known example. Upon D'Alembert's
surprise at the answer "Why not?", Diderot develops this process of
transformation:

> D'Alembert: "I'd like you to tell me what difference there is according to you between
> a man and a statue, between marble and flesh.
> Diderot: Not much. Flesh can be made from marble, and marble from flesh. . . . It
> occurs every time you eat.
> D'Alembert: Every time I eat?
> Diderot: Yes, for what do you do when you eat? You remove obstacles that pre-
> vented the food from having an active sensibility. You assimilate it, you turn it into
> flesh, you make it animal, you give it the faculty of sensation, and what you do to this
> foodstuff, I can do, when I please, to marble.
> D'Alembert: And how?
> Diderot: How? I shall make it edible.
> D'Alembert: Make marble edible? It doesn't seem easy to me.
> Diderot: It's up to me to show you the process. I take the statue you see there, I put
> it in a mortar, then with heavy blows from a pestle [I beat it into a powder]. . . . When
> the block of marble is reduced to unpalpable powder, I mix it with pumice or leaf
> mold; I knead them well together; I water the mixture, I let it decompose for a year, or
> two, or a hundred, time doesn't matter to me. When the whole has turned into a more
> or less homogeneous substance . . . you know what I do?
> D'Alembert: I'm sure you don't eat the humus.
> Diderot: No, but there is a means of connection, of assimilation, a link between the
> humus and myself, a *latus* as the chemist would say.
> D'Alembert: And that is plant life?
> Diderot: That's right. I sow peas, beans, cabbage, and other vegetables; these plants
> feed on the soil, and I feed on the plants. . . . Here in four words you have the general
> formula: Eat, digest, distil *in vasi licite et fiat homo secundum artem*. And to expound
> before the Academy the process of the formation of a man or of an animal, one need
> employ only material agents, the successive results of which are an inert being, a feeling
> being, a thinking being. . . . [A.-T., II, 105–110]

Clearly, it is not the stone as stone that feels. It is not matter as matter
that is sensible. But matter is capable of sensation and thought. It is this
sense of the potentiality in matter of having sensibility, that has led to the
misinterpretation of Diderot as a vitalist. Though Diderot sometimes
speaks of sensibility as a general property of matter, it is only in this
sense of process or development. He flatly denies the existence of
pre-existent prototypes (A.-T., II, 110), which occur in Robinet,
Maupertuis, and La Mettrie, as 'germs,' 'animalcules,' 'semences.'

This transformism or process of change is a temporal development of matter in motion; time, then, is an objective condition for the phenomenon of change. The instantaneous mutation of the scholastics and early mechanists is resolved by Diderot by making temporality a materialistic mode of the existence of matter. The process of making marble into flesh, which takes place by 'purely mechanical means,' is not a process in terms of mechanistic local motion or change of place. Remembering the identity of force and motion in Diderot, we can see what kind of mechanical process is involved here. He speaks of animate and inanimate force, of the *molécule morte* and the *molécule vivante*. Inanimate force does not mean force without motion, as was argued by Maupertuis in the *Essai sur la cosmologie*, and does not mean matter at rest. Inanimate force is potential force or motion, the potential expression of the force inherent in the molecule. Animate force is the actual expression of force in translation or local motion, in action of the molecule outside of itself. Diderot says that

. . . the transference of a body from one place to another is not motion. It is the consequence of motion. Motion exists equally in the body displaced and in the body that remains stationary (A.-T., II, 106).

In this sense, sensibility is potential in matter, insofar as it is the product or the result of motion. In a scientific speculation as to the relation of motion to energy, Diderot relates motion to heat, as a means of explaining the rise of sensibility in matter:

How does this insensible mass pass to another level of organization, to sensibility, to life? By heat. What produces heat? Motion (A.-T., II, 115).

The change from inorganic to organic matter is achieved in a sense 'mechanically,' as a result of motion giving rise to heat. The mechanical motion of change of place is only the appearance, not the essence of this process of change. Its essence is in the fact of motion itself, of which change of place is but a result. Therefore, change of place is only one side of this transformation. It is the organization of self-motive matter in a certain way, under certain conditions, that gives rise to organic life. Diderot comes remarkably close to the modern scientific theory of what this particular organization is, in the *Rêve de d'Alembert*, where he speaks of the germinal substance of organism as being 'filaments' of a gelatinous substance, and of chromosome-like 'bundles of filaments' as the bearers of the continuity of species. With a materialistic explanation

of the process of transformation that takes place from egg to chick, Diderot claims that all the miraculous explanations of the origin of life will collapse: "Do you see this egg?" he asks, "With this egg I am ready to overthrow all the schools of theology and all the temples of the earth" (A.-T., II, 115).

This change from inorganic to organic matter is, for Diderot, a change in the qualitative level of the organization of matter. These qualitative differentiations within the monistic chain of being characterize his monistic materialism. The aggregates that he speaks of are not merely quantitative combinations, but are qualitative levels of the organization of matter. Thus continuity and discontinuity, the unity of particularity and universality, of quantity and quality, are maintained by Diderot as characteristics of matter in motion. This is not a simple metaphysical unity, not an absolute subsuming of opposites such as we find in the celestial realm of scholasticism, or in Leibniz's monad where there is a metaphysical unity of opposites, or in the metaphysical dialectic of Schelling, but it has the characteristics of such a unity of opposites where the opposition is not merely negated or ignored, but where the very condition of the unity itself is opposition. The pre-Hegelian dialectical element is based on the essential role of process, dynamism, development. The levels are the product of a process in matter, are not preordained, are not prototypes. The flux in Diderot's universe is not a flux-in-itself, it is a flux grounded in matter, in the mode of the existence of matter: motion. Although he never systematizes this process in philosophic terms, it is an essential element in his transformism.

The swarm of bees in the *Rêve de d'Alembert* serves as an example of an aggregate of matter, both quantitatively and, under certain conditions, as a distinct qualitative level. This cluster of bees is a continuity in the sense that, as a distinct being, it has its own characteristic 'motion,' its own development. But, says Diderot, the swarm of bees is as close as we can get to the *continuity* characteristic of an organism, only if we 'glue together' the feet of the bees; otherwise it is only contiguous. In the case of the aggregate which is continuous, there is an internal interconnection, not the unity of an atomistic universe, but a real organic unity subsuming the distinct particulars. Between the elements of this unity, there is continual action and reaction, a constant internal process going on. It is not a static, but a dynamic unity. The continuity or relative continuity of the aggregate makes it a qualitative level of the organization of matter, not merely a collection. For Diderot,

contiguity and continuity are related. A polypous animal, such as a worm, is a homogeneous aggregate, which can, when cut up, go on existing as two separate continuities. Man, says D'Alembert in the *Rêve*, is a continuity of a different, more complex sort:

All our organs . . . are separate animals held together by the law of continuity in a general sympathy, unity, and identity (A.-T., II, 127).

But no design or final cause gives rise to these forms or aggregates, as in Maupertuis. At the level of organic matter, these changes take place in a pre-Lamarckian sort of evolution, which was a current theory derived in part from Lucretius and in part from the studies of comparative anatomy and geologic evolution by Buffon, whose most brilliant pupil was Lamarck. This is expressed by La Mettrie in the form: "Organs produce needs, and reciprocally, needs produce organs" (*Oeuvres phil.*, I, li). The renewed interest in Lucretius, stimulated by Buffon's work and by the mounting pressure for social change, which was transferred into other than social or political contexts, led to a fairly common agreement among the French materialists on the evolution of forms. Naigeon, editor and close friend of Diderot, had just published a new translation of *De Rerum Natura*, and evolution was the order of the day for scientific speculation. Change, which in the scholastic and Aristotelian traditions, was a principle of mortality, of self-destruction, of imperfection, was, in French materialism, united with its opposite, the eternal substance, the perfect existence. Change was seen by Diderot as a phenomenon organic to eternal matter. Change, like motion with which it is identified, is heterogeneous, particularistic, both quantitatively and qualitatively. Nature perfects itself, eliminates 'monsters,' 'vicious beings,' in its own inherent process of change. But throughout, change means change of *something*, change of a whole, which itself is infinite and eternal. La Mettrie writes in the *Système d'Epicure:* "Everything succeeds itself, everything disappears, and nothing perishes" (*Oeuvres phil.*, I. 61). There is not merely change, there is development. The change is from lower to higher forms, and for La Mettrie, the instrument of this change is an embryonic survival of the fittest, which we shall also see in Diderot.

The first generations must have been quite imperfect. . . . It is evident that the first animals which were able to survive and to conserve their species would have been those which found themselves equipped with all the parts necessary to generation and from which . . . no essential part was missing. . . . Perfection couldn't have been the work of a day in Nature, any more than it can in Art. [*Oeuvres phil.* I, xiii]

Animals, says La Mettrie, are less perfect than man, perhaps because they were formed earlier. Man himself is nothing but a refined sort of animal, he maintains, ". . . since the entire realm of man, to tell the truth, is nothing but a composite of different monkeys at whose head Pope has placed Newton" (*Oeuvres phil.*, I, xxxiii).

La Mettrie also speaks of a geologic evolution in which water originally covered the earth's surface, was absorbed, leaving the eggs of different species lying in the moist earth; and the sun, acting as an incubator, hatched out these eggs and brought forth the organisms potential within them. The forms all seem to pre-exist somehow in these original eggs. But the early forms are not the ones we know today. These early forms hatched out of the eggs are the raw stuff of La Mettrie's theory of evolution; they intermix, the more perfect results continue their species, the "monsters and vicious forms" die out in this continual hybridization of the original forms. Thus, in one of La Mettrie's wilder speculations, the different 'species' of men are the results of the crosses between men and various animals.[10]

In the early *Lettre sur les aveugles*, Diderot already speaks of an evolutionary development of forms and worlds. The blind English mathematician, Saunderson, says to the priest who has come to give him the last rites:

You may imagine, if you want to, that the present order with which you are so much impressed, always subsisted; but let *me* believe that it hasn't, and if we were to go back to the birth of things and of eras, and if we perceived matter in self-motion, and the fog clearing away from the chaos, we would encounter a multitude of unformed beings for each well-organized being. . . . The monsters destroyed themselves successively, all the vicious combinations of matter disappeared . . . only those survived whose mechanism did not have any important weaknesses or contradictions, and which were able to exist and to perpetuate themselves. I speculate then, that at the beginning, when matter in fermentation gave rise to the universe, my species [the blind] were quite common. But why cannot I also posit for worlds what I believe about animals? How many atrophied worlds are missing, have dissipated, are reforming and dissipating themselves, perhaps at each instant, in the distant spaces, where motion continues and will continue to combine masses of matter until they become arranged in such a way that they can persevere. . . . What is this world, Mr. Holmes? A composite, subject to revolutions, which indicate a continual tendency towards destruction; a rapid succession of beings which follow upon each other; an ephemeral symmetry, a momentary order . . . the earth is eternal for you as you are eternal for the being which is sensitive for only a moment. [A.-T., I, 309–311]

This early statement is reflected in the *Eléments de physiologie*, where

Diderot speaks of the extermination of 'contradictory beings' by a blind nature, because they 'cannot co-exist in the general order.' Of the species of animals, he writes: "One mustn't believe that they have always been and will always be the way we see them" (A.-T., IX, 264).

This evolutionary change includes all of being, not merely organic being. All the lower and higher forms develop from the original heterogeneous matter:

The vegetable realm could well have been the first source of the animal realm and could have had its origin in the mineral realm; and the latter could have arisen out of universal heterogeneous matter (A.-T., IX, 265).

All the forms are transitory, but the totality, heterogeneous matter in motion, is eternal:

The general order is constantly changing; how in the midst of this change can a species endure in the same form? Only the molecule alone remains eternal and unalterable (A.-T., IX, 418).

The changes in organic species, Diderot says, are brought about by the reciprocal action of organs and needs; he develops the theme suggested in Epicurus and stated by La Mettrie in his *Système d'Epicure:*

The organization determines the functions and the needs, and these can sometimes go as far as to produce new organs, and always go as far as to transform them.

The pre-Lamarckian correlation between needs and the transformation of organs indicates that for Diderot this transformation is not accidental. The fittest survive, the 'monsters' are destroyed; but this adaptation by elimination pertains not only to organisms as such, but also to organs, to parts, and to organic 'instruments.' Thus the evolution of species is not a chance one, but rather, needs actually bring about causally a transformation of the species, as the organic structure of the species brings about a transformation in its needs, an adaptation to, and an adapting of, an environment. Diderot offers only the most general sort of scientific evidence for this phenomenon, only the observation of the seeming suitedness of organisms to their environment; but his rejection of teleology leads him to an advanced, though speculative formulation of evolutionary change. He describes this constant universal change thus, in the *Rêve de d'Alembert:*

Who knows what species of animals preceded us? Who knows what species of animals will follow ours? Everything changes, everything passes away, only the whole remains. The world is forever beginning and ending. At every instant it is at its beginning and at

its end. It has never had and never will have any other beginning or end. In this vast
ocean of matter, no Molecule is for one moment like itself. *Rerum Novus Nascitur Ordo*
is eternally inscribed on it. [A.-T., II, 132]

Change itself is eternal. It is coexistent with, or rather, *is* the existence of
matter. Matter, motion and change, in this relation of inner identity, are
the cornerstones of Diderot's materialist monism. This process of
change, of qualitative transformation through the interaction of needs
and organisms is stated in the *Rêve de d'Alembert* thus:

Dr. Bordeau: Assume a long succession of armless generations, assume continual
efforts, and you will see the two ends of this pincer stretch further and further, cross at
the back, come round in the front, perhaps develop fingers at their ends, and make arms
and hands once more. The original conformation degenerates or is perfected by necessity
and by normal function. We walk so little, we work so little, that I don't despair that
man will end by being only a head.

Mlle. de L'Espinasse: A head! A head! That's not very much. I hope that excessive
lovemaking won't . . . but you're suggesting some very ridiculous ideas to me. . . .
[A.-T., II, 138]

And in the *Eléments de physiologie*, Diderot similarly writes: "Nature
adapts itself to function. I am not convinced that the long suppression of
an arm might not lead to an armless race" (A.-T., IX, 419).

The interconnection of organ and environment, of need and
transformation, is, at the level of organic matter, the same total internal
and external interconnection of the molecules: the interrelation of
qualitative levels, of aggregates, is the basis for Diderot's monism at the
level of aggregate, just as this interrelation is at the basis of his monism,
at the level of molecules, of elements.

D'Alembert: So I am what I am because I had to be so; change the whole and you will
necessarily change me; but the whole is constantly changing. . . . Man is merely a
common product, the monster an uncommon product; both equally natural, equally
necessary, equally part of the universal and general order of things . . . and what is so
astonishing about that? . . . All creatures intermingle with each other, all species . . .
everything, is in a perpetual flux . . . every animal is more or less man; every mineral is
more or less plant; every plant, more or less animal; nothing is precise in nature.
[A.-T., II, 138-139]

Diderot, continuing this passage, harks back to an element contradictory
to his theory of the development of sensibility in matter as a product of
organization and formulates a theory similar to Maupertuis', in which all
qualities inhere more or less in all matter, and the difference in forms is
only a difference in quantitative degrees:

. . . There is no quality of which no being has a share . . . and . . . it is the greater and less degree of this quality that makes us attribute it to one being to the exclusion of another [A.-T., II, 139].

Further on in this passage, Diderot has D'Alembert follow Maupertuis almost entirely, in speaking of the absence of individuals and of the existence of one individual: the whole. Furthermore, he repeats the assertion that in a qualitatively distinct being all the inherent perceptions of the parts are subsumed in the perception of the whole. But the sense of this identity of qualities in all organisms has a deeper element than Maupertuis' atomism of the perceptive, elementary particles, though it is stated unclearly in this instance: that is, the lack of individuality of the forms is the lack of absolute, essential, metaphysical individuality, like that of the impenetrable atoms. The passage is directed against the imputation of essences to matter by "poor philosophers." It is directed against essences as absolute, isolated, real entities, outside of their contextual existence. For Diderot certainly attributes essential character-istics to matter and to aggregates of matter. But these are essences or qualities only *in* their total interconnection, only as temporal functions of a whole in process. Diderot concludes that there are no fast, hard lines in nature, that there is an infinite gradation of species, that nature has no absolute, eternal qualitative levels, but that qualitative levels are the product of a process of the whole of the material universe. Life, being a real quality, is arrived at by action and reaction and by quantitative changes:

. . . Life is a succession of actions and reactions. . . . Living, I act and react as a mass . . . dead, I act and react as molecules . . . I don't die then? No, doubtlessly, I don't die in that sense . . . to live, to be born, to pass away, is to change forms . . . (A.-T., II, 139–140).

Though not fully developed or systematized here, Diderot's approach is a dynamic one, superseding the mechanist theory of merely quantitative change. Diderot's constant concern with the paradox, with the contradiction, with the unity of a particularistic, qualitatively-defined chain of being, marks, within French eighteenth-century materialism, the first steps in the rejection of materialist mechanism and of vitalism. It means the formulation of a materialism without the idealist vestigial elements in vitalism and without the idealism that mechanism, out of its insufficiency, has to let in through the back door. Any analysis of Diderot which does not grasp his sense of internal opposition and contradiction,

his embryonic dialectic logic, must reveal him as a combination of two one-sided points of view, irremediably contradictory, or at best, with Cassirer, as constantly shifting his point of view, skipping about nonchalantly from peak to peak, in order to get a whole view. Nevertheless, this contrariness of Diderot's, in seeming at one point to state unity, continuity, eternality, and at another to state the opposites, particularity, discontinuity, temporality, is a statement, not of a subjective change of viewpoint, but of an objective contradictoriness in nature, in matter itself.

The role of the classification of species in natural science is not obviated for Diderot by the unity of nature. It is the first and necessary step which prepares the ground for scientifically establishing the interconnection between the different species and qualitative levels which itself leads to the realization of the unity of nature. The distinctions and gradations of classificatory science are not less real than the unity of nature.

It is necessary to begin by classifying beings, from the inert molecule, if there is one, to the active molecule, to the microscopic animal, to the animal plant, to the animal, to man. . . . One mustn't believe the chain of being to be interrupted by the diversity of forms; the form is often only a mask which deceives, and a missing link exists perhaps in an unknown being which the progress of comparative anatomy has not yet been able to assign to its true place. [A.-T., IX, 255]

The continuity of the gradations of being had been philosophically discussed by La Mettrie in *L'Homme plante,* by Robinet, in *Considérations philosophiques sur la gradation naturelle des formes de l'être,* and by Buffon, in *Epoques de la nature.* But La Mettrie for one, in the consideration of the relation of plants to animal life, had not developed a causal dynamic interrelationship between the two realms, but in the mechanist limits of classificatory science, had dealt merely with the comparison of the two to show the uniformity of nature. Nevertheless, he attacked those who considered their classifications as absolute isolates:

We begin to see into the uniformity of nature . . . this ladder [of being] so imperceptibly graduated, where one sees nature passing exactly through all its degrees, without ever jumping in any way a single rung, in all its diverse productions. . . . For myself, interested only in philosophy, an active nature will always be my sole point of view.[11]

For La Mettrie, the interconnection between plants and animals is a static, comparative one, rather than a causal one. The leaves are

compared to lungs; roots and capillaries find their animal counterparts. For Diderot, however, the organic causal interrelation between plant and animal is basic to his monism. The errors of the method of classification, the method of exclusion of non-essential qualities, as in Linnaeus, is criticized. By this method, for example, man is defined as a quadruped. Thus, "when following a wrong road, the faster one walks, the faster one goes astray," comments Diderot, in his *Pensées sur l'interprétation de la nature*. The organic process of the interrelation of mineral, plant, animal, is given in the example of the edible statue. The dynamic of this causality between qualitative levels of matter is an internal dynamic, a cause and effect, not from some first motion, but a function of the very nature of matter. What is this function, then, in the aggregate-animal? It is an activity of the aggregate which brings about a unity of plant and animal. The dynamic of this unity, eating, digesting, transforming food into animal tissue (the internal interconnections at the level of organism), is analyzed in a thoroughly materialistic way. "Every animal is more or less plant, every plant more or less mineral or animal," then, only in this sense of an active causal interrelation between the realms: "By heat and fermentation, vegetal matter becomes animalized . . . This happens in me . . ." (A.-T., IX, 256).

The relation of cause and effect at the level of organic matter is no more a mechanical cause and effect than the cause and effect relation between the molecules discussed previously. The cause and effect is a function of the internal properties of the organic aggregates here, just as it is a function of the internal force and motion of the molecules. It is this kind of cause and effect, this kind of determinism that transcends the quantitative limits of mechanism and gives rise to Diderot's transformism. The actions and reactions of organic matter, the interrelation of organs and needs, form a dynamic cause and effect which can explain the phenomenon of qualitative change in a fully materialistic way. When asked which came first, the chicken or the egg, Diderot resolves the metaphysical isolatedness of the two categories presupposed in the question, through his dynamic interrelationalism. Chicken and egg are for him active changing forms, not static prototypes:

If the question of the priority of the egg over the chicken, or the chicken over the egg, embarrasses you, it is because you suppose that animals were originally what they are at present. What folly! (A.-T., II, 110)

The ultimate answer to this apparent paradox is for Diderot *activity, process, change* itself.

This embryonic theory of evolution is thoroughly materialistic in that it admits of no ideal, non-material processes or entities. Soul is eliminated from the reasoning about qualitative differentiation in being:

The soul is a principle of life, the knowledge of which is not attainable without the wings of theology. And with these pretty little bat-wings, to what does one raise oneself? To nothing; one circles in the darkness (A.-T., II, 298).

The rise of qualitative forms in nature takes place out of the inherent flux of the universe itself. Quantitative change becomes transformed into qualitative change, without either design or divine intervention, but as a characteristic of matter itself. The continuum of forms, the preservation of species is in itself only relative. The germ-plasm, the 'gelatinous mass' or substance which carries within it the potentiality of the reproduction of species, is not itself immortal or unchanging, but itself participates in the constant action and reaction, in the complete determinism of Diderot's dynamic monism, itself becomes modified and transformed. It is on the subject of this early, embryonic concept of germ-plasm, that Diderot makes advanced scientific speculation. He develops a concept of chromosomal structure and activity very like that revealed by later discovery. Diderot was steeped in the contemporary chemical researches into the composition of the elemental living substance, as passages in the *Eléments de physiologie* and the *Rêve de d'Alembert* show. In the former work, he reveals an intimate acquaintance with contemporary research, derived in part from his studies with the chemist Rouelle, Lavoisier's teacher. Diderot comments on the discovery of gluten by Beccari in 1742. This gluten seemed to be a sort of animal substance. Dumas and Cahours defined glutenous stuff as being vegetal-fibrinous. Rouelle and Macquer defined gluten as vegetal-animal. This gluten, which seemed to defy classification as animal or plant, pointed up for Diderot the possibility of a common organic stuff, which, being itself a level of material organization, had as its forms vegetal and animal matter. It pointed up also the absence of a hard and fast line between animal and plant forms. Diderot gives as an example of such a problematic link, which defies clear classification as animal or plant, *Adamson's Tremella*. The glutenous, protein-like, fibrinous substance, by its different combinations and organizations, gives rise to different forms; that is, this plasm itself becomes qualitatively modified, and these modifications have a relative perseverance and duration which distinguishes species. This sort of plasm becomes the carrier of hereditary characteristics in Diderot's

analysis. The statement of this in the *Eléments de physiologie* is oversimple and reminiscent of Maupertuis' *animalcules*, in themselves the man, the dog, the elephant, wholly formed, wholly potential miniatures growing equally outward. But these *animalcules*, even here, are not pre-existent for Diderot, but are products of the previous animal: that is, the whole animal substance has a hereditary potentiality. The development of the idea of a specialized hereditary substance made up of this 'fibrinous, soft, elastic, gelatinous' stuff is found in the *Rêve de d'Alembert*. At first, Diderot has D'Alembert speculate in his dream of the preservation of polypous homogeneous species:

On Jupiter, or on Saturn, human polypi! Males splitting up into males, females into females. Man splitting up into an infinite number of atomic men that can be wrapped between sheets of paper, like insect eggs that spin their cocoons, remain as chrysalids for a certain time, then break through their cocoons and escape like butterflies. A society of men formed and a whole province populated out of the fragments of a single man. [A.-T., II, 130]

He then speculates on this polypous, homogeneous type of heredity passing on through the heterogeneous organs of the man-aggregate: "Doesn't the dissolution of different parts produce men of different characters? The brain, the heart, the chest, the feet, the hands, the testicles . . . oh! how this simplifies morality!" (A.-T., II, 131) All of these hereditary characteristics are in the first gelatinous 'speck' that constitutes the organism at its inception:

At first you are nothing at all. You began as an imperceptible speck formed from still smaller molecules, scattered through the blood and lymph in your father and mother; that speck becomes a loose thread, then a bundle of threads. Up until then, not the slightest trace of your own . . . form . . . . Each of the fibres in the bundle of threads was transformed solely by nutrition, and according to its conformation, into a particular organ; exception being made of those organs in which the fibers of the bundle are metamorphosed, and to which they give birth. [A.-T., II, 144]

This latter 'exception' seems to indicate a specialized organ in which the formation of this gelatinous substance, charged with heredity, takes place. Nevertheless, it is out of the different organs themselves, through the continuity and unity of their own aggregate, that primitive sorts of genes, called 'fibers,' are formed. The interconnection between the heterogeneous organic substances in the animal aggregate and the germ-plasm which contains them all is, at this level of organic animal being, another such unity of distinct qualities, interacting as need and organ interact. The origin of the different 'fibers' in the bundle of

threads is in the different organs themselves, according to Diderot. Furthermore, Diderot speculates, the suppression of one of these fibers would lead to the underdevelopment or total lack of development of the organ or part it was to form: "Do in your mind what nature sometimes does actually; deprive the bundle of one of its fibers, for instance, of the fiber which should form the eyes; what do you think will happen?" (A.-T., II, 147) The animal, the speculation continues, would have no eyes, or one eye. By duplicating these fibers, on the other hand, the animal would have four or six eyes. But it is not only the particular organ which would be deformed, since every particle, every action of a part of the whole is interconnected with every other particle or action in

an organism whose regular or irregular structure depends on a bundle of thin, loose, flexible fibers, a sort of skein in which the slightest fiber cannot be broken, snapped, displaced, or removed, without distressing consequences for the whole . . . (A.-T., II, 149).

This is recognized as pure speculation by Diderot, but as scientific speculation based on material explanation, accessible to scientific proof. Dr. Bordeu comments on this speculation thus:

That was a fine flight he made; that was very lofty philosophy; only theoretical at the moment, yet I believe the more progress is made in human knowledge, the more will its truth be confirmed (A.-T., II, 140).

This entire theory of the passing on of hereditary characteristics and the disturbing of the genetic stuff by 'suppression and entangling' of the fibers serves for Diderot to explain both the preservation of species and the variety in forms within species. In this sense, the origin of 'monsters' is as necessary as that of normal beings. When this genetic continuity is combined with the theory of the causal interconnection of organ and need, organization and activity, then the hereditary stuff itself is no longer an immortal, preexistent *animalcule*, but is somehow interconnected with the whole activity and change in the organism.

Diderot's further developments of this monist materialistic approach with respect to consciousness, thought, and reason, need not be dealt with here. His original contribution is the development of a philosophic monism which places man and all nature on the solid ground of a material universe. Our knowledge, our ability to cope with this universe scientifically is assured by the very unity of our consciousness or reason with the material world, by the fact that our consciousness has its very source in the qualitative development of matter. This new realization of

the materiality of consciousness, and of the power of thought to effect change in the material world, is an important aspect of the Age of Enlightenment.

NOTES

[1] Denis Diderot, *Oeuvres complètes de Diderot*, ed. Assézat-Tourneux, (Paris, 1875–77). All references to this source are indicated by "A.-T.," indicating the edition, followed by the volume and page(s) referred to.

[2] Isidore Sherman, in his study on Diderot, separates what he calls the 'metaphysical thesis' (materialism) and the 'naturalistic thesis' (monism), a rather sorry dissection of what, in Diderot, is a unity.

"The 'metaphysical thesis' is simply a corollary of the materialist thesis: If reality is made up of any one substance then the universe and everything in the universe constitutes a unity, an unbroken continuum. But on another interpretation, *the monistic thesis can be taken as an independent axiom apart from any commitments to the ultimate nature of reality*. (My italics.) In this latter sense I shall refer to it as the 'naturalistic thesis' – the monistic thesis is simply a statement about certain relations between different parts of nature, no matter what the ultimate stuff of these parts may be." (Isidore Sherman, 'Diderot and Philosophic Naturalism,' Master's Essay, Columbia University, 1945, p. 9).

[3] The question of Diderot's early deism is debatable. We are of the opinion that by the time he wrote the *Pensées philosophiques* his deism was a disguised atheism for the benefit of the royal censor, a tongue-in-cheek respectability belied by the content of the 'deist' writings. For an interesting argument for Diderot's deism see 'From Deist to Atheist – Diderot's Philosophic Orientation' by Aram Vartanian, in *Diderot Studies*, ed. Fellows and Torrey, (Syracuse University Press, 1949), pp. 46–63.

[4] Feuerbach's comment is enlightening here: "Spinoza is the negation of theology on the basis of theology."

[5] Maupertuis, *Oeuvres* II, Lettre viii. (Lyon, 1768), 3 vol. All subsequent references are to this source. References to the *Système de la nature* are indicated by the volume in the *Oeuvres*, and the abbreviation "Système," followed by the author's section number in lower case Latin numerals.

[6] Ernst Cassirer, *Philosophie der Aufklärung* (Tübingen, 1932), p. 116.

[7] I.K. Luppol. *Diderot – Ses Idées philosophiques*, translated from the Russian by V. and Y. Feldman, (Editions Sociales Internationales, Paris, 1936), p. 233.

[8] Robinet. *Système de la nature*, cited by Hegel, *Lectures on the History of Philosophy*, translated by Haldane and Simson, III, p. 393.

[9] La Mettrie, *Oeuvres philosophiques* (Amsterdam, 1764), 3 vol. I, vi. The references to the *Système d'Epicure* in this edition are indicated by the volume and the author's section number in lower case Latin numerals.

[10] Cf. Diderot, *Satire I*, (*Sur . . . les mots de caractère. . . .*): "[Our reason] . . . corresponds to the instincts of all the variety of animals . . . there is the man-wolf, the man-tiger, the man-fox . . .", for Diderot's light comment on this idea. (A.-T., VI, 302).

[11] La Mettrie, *L'Homme plante*, republ. with Introd. and notes by L. Rougier; Publications of the Institute of French Studies, Columbia University, (New York, 1936), pp. 116, 143, 146.

ART AND TECHNOLOGY: CONFLICTING MODELS OF EDUCATION? THE USES OF A CULTURAL MYTH

[1973]

There is a current and popular distinction made between art and technology as conflicting modes of human activity. Art, as we all know, is creative, liberating, free activity. Technology, as we all know, is mechanical, constraining, subject to rote and rule. Where art calls for originality, invention, the fullest play of the imagination, technology is no more than acquired skill, transmitted by training, requiring at most imitation, or an ability to follow instructions. Such a popular characterization takes on mythic proportions: the alternative modes easily become 'models,' emulated and mapped onto various domains. We recognize the distinction, or its analogy, in talk of alternative or conflicting 'life styles': the liberated and the repressed. We see its reflection in the contrast between play and work. In social contexts, the counterposition becomes one between a privileged elite, able to exercise its leisure creatively, and a mass, condemned to the workaday repetition of dead tasks merely instrumental, or, as 'alienated labor,' required by the system. Inevitably, the distinction comes to be mapped onto education, where art and technology may serve as conflicting models of the educational life and process as well.

This crude and current myth has little to recommend it; yet it is pervasive, accords with a common understanding and a prevalent mood, and perhaps worst of all, permits us to build theories of education which appear to be on the side of God, the angels and poor little Johnny. Who, after all, will *not* want children to have the opportunity to be 'creative,' 'liberated,' 'free' from the automatism of the factory system (whether in industry or in education), from the tyranny of rule and rote, from the dead hand of mere artisanal skill? The technological model of education, so the myth suggests, is dedicated to the production of cogs for the machine. The 'cogs' are well-fitted and well-trained products needed to keep the system going. The schools are therefore nothing but the factories and assembly lines for such social production, and their ends are dictated not by the needs of children, for growth and development as

human beings, but by the needs of the system, for craftsmen and functionaries. Much follows, on this model, concerning the details of education: the role and function of the teacher; the requirement and rationale for discipline; the shape and aims of the curriculum. The force of a model as metaphorically rich as the technological one, is that it suggests so clearly how all these details fit into a coherent pattern. And it permits – even encourages – us to think that we can understand and explain the phenomena at issue – the crisis in education, the anomie of the student, the frustrations of the humanist-teacher, the vagaries of funding.

   As with any myth, there is enough of the truth in its stark formulation of alternatives to evoke in us an assent which draws its force from what we add from our own and varied experience to flesh out the details of the dramatic sketch. Surely art is a liberating and exhilarating experience, in which the highest powers of human imagination and creativity find expression. Certainly art is an exaltation of the human spirit, an intensification of everyday experience, a fit activity for free men and women, and therefore a fit model for the education of children. Just as surely, the image of the educational system in the bleak and utilitarian colors of a training ground for the industrial and technological army has a notable veracity. The school as factory, as ideological preparation for service to the industrial-capitalist system, as inculcator of the proper habits and attitudes required by 'the system' has been the object of a long and devastating critique, as well as of more and less successful efforts at humanistic and liberal reform. That mass public education as a legal requirement grew out of the needs of the industrial revolution is a patent fact. It is nowhere more vividly described than in Karl Marx's chapter on the Factory Acts in England, and on the institution of compulsory education for working children as part of the Child-Labor laws.

   What is wrong with the myth, however, is not that it somehow embodies an enlightened attitude toward the aesthetic and a critical one toward technology, but rather that if falsifies both art and technology, misunderstands the very activities which it takes as its abstract models, and ends with a shallow and comfortable humanism where what is needed is a deep-going and hard-headed radicalism.

   Therefore, I want to make a case against this model of an antithesis between art and technology, and specifically, against its interpretations in educational theory and practice. I will suggest further that there is much that is culturally and educationally vicious in this confrontation of the two modes. Moreover, I will argue that it is politically vicious as well,

and that this counter-position of art and technology hides, beneath its
rosy humanist appearances, a bourgeois-liberal myth, whose effect, like
that of Plato's Royal Lie, in the *Republic*, is to keep the working masses
in their place. My argument is somewhat perverse, for it will *appear*, in
the context of the myth at least, that I am arguing for repression, against
liberation of the free spirit of the imagination, and for rote and rule. In
fact, I am arguing against the myth, against that view of art and the
concomitant view of technology, which yields this counterposition in the
first place. And I want to show that the conceptions of 'art' and
'technology,' in this mythic form, are sham conceptions, and that their
perpetuation is part of a self-serving 'liberal' ideology; that this ideology
(which often pretends to the title 'radical') simply seeks to cast education
in the image of allegedly enlightened middle-class values, to the
denigration of working-class values; and that its effect, were it
successful, would be to plaster over the worst failures of our
mass-education of the young with a fig leaf to hide our shame. Now it is
clearly *not* my intention to denigrate either art in the schools, or the
richness of an aesthetic model of education. It *is* my intention to rescue
the technological from the elitist disdain in which it has been held by
often well-meaning liberal and humanist social and educational theorists.
The confusion is clear: 'technology' has been identified with the abuses
of a technological-scientific society, with the misuse of rationality in the
service of exploitation and war. This is the superficial Luddite critique of
technology: smash the machines, get back to the soil and the sunshine,
and all will be well. The deeper critique of technology argues more
sophisticatedly that the very conditions of social organization and work
made necessary by a rationalized technology are inherently brutalizing,
anomic, alienating; that the essential requirements of technological
advance inevitably lead to pollution, ecological disaster in the rape of
resources to feed the machine, and a mechanized and dehumanized form
of human labor. Both versions, however, offer no solution short of
deindustrialization, or the more vicious alternative of 'zero-growth.'
More to the point, in the popular forms in which antitechnology is
disseminated, technology is seen as no more than the soulless, amoral,
and value-free means to an end, whereas ends themselves come,
somehow, from beyond the 'merely technological,' i.e., from a
'value-system' or from the 'culture.' The fact-value dichotomy is neatly
packaged so that human judgment concerning ends is abstracted from
the concrete contexts in which choices become practical and actual:

within the constraints of actual needs and of actual means to satisfy them. Once the separation is effected, the ease with which a rosy humanism can generate ends at its pleasure and leisure becomes boundless. Utopianism is the name of the game. Pie in the sky is the name of the payoff. The masses of the poor and the underprivileged suffer because utopias can be realized only in colonies, in 'experiments,' in 'model-programs,' which filter down only into the reports of committees, or into the glossy paperbacks which serve dreamy educators in much the same way that 'True Romances' or 'Muscle Power' served the dreamy youth of an earlier generation. The teacher, surrounded with broken desks, deprived and disturbed children, inadequate supplies, an impossibly tough and retrograde school committee, and proliferation of blindly bureaucratic paperwork, dreams of the 'child-centered school,' of the 'authentic educational encounter,' even of Room 222 with its inane mix of happy children with happy problems, cutesie teachers, and a Jewish Mother principal. My point is that the ends separated from the means (political, technological, practical, humdrum, organizational) become shallow and effete; and that the antitechnological model of technology serves in practice to effect this separation.

I'm exaggerating for effect, and being deliberately cruel in this characterization. For, in effect, the specter of technology which haunts the bourgeois-liberal theory of education needs desperately to be laid to rest. A specter, we know, is a construction of the imagination; thus, we are able to invest it with our fears, anxieties, and hatreds. Specters *are* embodied fears, after all. But the specter of 'technology' has been invested with the evils of education, much as has the goat which bears our sins for us. The dulling and deadening characteristics of mass-education, the packaged lesson plan, the programmed sequence which determines which of our cortical centres shall fire when, the 'transmission' of 'knowledge,' as a stuffing operation, by the book and by the clock, the bureaucratization of the school systems, the standardization of 'progress' in learning, the lockstep grade-system, the uses of the IQ, all have the earmarks of 'technology': the school as assembly line; the teacher as a one-operation mechanic, coming in at the appropriate year, or semester, or hour, and then leaving; the student as product – all of these very real problems conduce to the characterization of the technological in these terms, as the source and 'model' of our educational ills.

Enter art. Here, not the constraints of necessity, but the air of freedom reigns. The feelings expressed, the mind open to inspiration, invention, discovery; the unique and the individual in each of us fully encouraged. Gone, the serried ranks, gone the knee-jerk responses of obedient little minds, gone the barracks atmosphere. Air, light, and sun, O art! O freedom!

This vision of art as a liberation from the constraints of a mechanical necessity, as a corrective to the very real problems of a bureaucratized and inhumane educational practice, fails not merely on the grounds of its utopianism, but fails as a vision of art. Art as an escape from drudgery and necessity is escapist art. It changes nothing. It accommodates the present situation and its oppressive tendencies. But beyond this, it falsifies the very nature of that disciplined creativity which art requires, that necessity of craft and skill, that training without which 'art' becomes a sham.

The artist – the real, live working artist – is no less, and often more a 'technologist' than any assembly-line mechanic or production lathe operator. He has to be, for the expression of his art requires the discipline of his craft. But perhaps the term 'technologist' grates here. We think of technologists as engineers, perhaps, or as computer technicians. They occupy a middle range between 'scientists' – i.e., theoreticians, creative types who 'use their minds' – and mere 'laborers' who use only their muscles, with whatever minimal skill that is required. The 'technologist' somehow effects what the scientist thinks out, but he has to use his mind little, and under supervision and direction. He is, at least, a high-school graduate, but not an academic. The internal class and status differentiation, while interesting, is beside the point here. On the model of 'technology,' as opposed to 'art,' it becomes the work of the world, the practical, and therefore constrained activity in pursuit of some tangible goal, in answer to some articulated need, and thus tied to earth, and more particularly, to the means for carrying out a prescribed end. In this sense, science too is easily misunderstood either as a theoretical tinkering or invention, or as purely theoretical 'art,' as an activity whose only constraints are aesthetic, or formal, which is therefore divorced from practice, or the conditions of application. But I want to use the term 'technologist' more broadly here, for the sake of the argument against the model which counterposes art to technology. The artist as 'technologist' in this sense is simply the artist aware of the limits of his medium, and therefore of its possibilities as

well; aware therefore also of the requirements for practical skill and
technical understanding in the service of his vision. To the extent that
work has been degraded so that vision and imagination play no part in it,
skill itself has sometimes been raised to the level of creative application.
The counterpart in art is virtuosity, as a quality abstracted from vision
and understanding. The worker, deprived of any role in the planning and
shaping of the product or the production process, will elaborate his skill
and take pride in it. When even this possibility is removed, and work
becomes rote performance of an unthinking and easily learned
procedure, then even the dignity of craftsmanship is robbed from work.
By analogy, when the imaginative refinement of skill and craftsmanship
is denied to the child, in educational contexts, he is likewise brutalized;
just as he is demeaned when skill itself is not informed with a vision
whose ends it subserves. But the model I am criticizing consigns skill and
craftsmanship to the domain of a denigrated technology, and conceives
of art, by contrast, in the image of an inane and empty 'freedom' and
'creativity,' in which wish becomes reality without the mediation of
work. The falseness of the separation works the other way as well: it
denies the aesthetic component of skilled performance itself. The child,
caught in the intricate web of an electrical circuit design, or in the
weighty determinations of fitting a tongue and groove, or of constructing
the proof of a geometrical theorem, has (or should have) as much
aesthetic experience in the situation as he has wielding a brush or
dancing. The fact that this is not the case in much actual school practice
reveals the extent to which technology itself has been subverted in the
interests of sheer rote learning for manpower needs. But the subversion
of the humane content of technological activity is not overcome by the
subversion of the technical requirements of aesthetic activity. The
humane content of technological activity is precisely that it is the activity
by which human beings produce the means of their existence. It is that
conscious praxis in which human teleology, the deliberate and conscious
organization of human skills toward a human end, is first developed
historically, and is renewed again as each generation transmits its
productive genius and achievements to the next. The dehumanization of
technology comes with the separation of the process from its ends, and
with the embodiment of this separation in exploitative class societies, in
which one person works for another's weal, and becomes no more than
an instrument for ends not his or her own. But the dehumanization of
technology calls not for its continued, and even sharpened separation

from 'free,' i.e. end-controlling, teleological acitivity, but for its *re*humanization. Here, the myth abdicates the struggle utterly. In its place we have a poultice misnamed 'art' or 'creativity,' in which the very idea of artistic creation has degenerated into some notion of the 'production' of instant art, of 'objects' requiring neither skill nor the delicious deliberation required for the difficult work of coordinating hand, eye, and heart.

So, not yet to the issue at hand, concerning 'alternative models of education,' but rather to the characterization (or mischaracterization) of art and technology themselves: Whence the distortion?

Santayana, in his discussion of work and play (in *The Sense of Beauty*), reconstructs the puritan myth for us. He writes:

We may call everything play which is useless activity, exercise that springs from the physiological impulse to discharge the energy which the exigencies of life have not called out. Work will then be all action that is necessary or useful for life. Evidently if work and play are thus objectively distinguished as useful and useless action, work is a eulogistic term and play a disparaging one. It would be better for us that all our energy should be turned to account, that none of it should be wasted in aimless motion. Play, in this sense, is a sign of imperfect adaptation. It is proper to childhood, when the body and mind are not yet fit to cope with the environment, but it is unseemly in manhood and pitiable in old age, because it marks an atrophy of human nature, and a failure to take hold of the opportunities of life. Play is thus essentially frivolous. . . . At the same time there is an undeniable propriety in calling all the liberal and imaginative activities of man play, because they are spontaneous, and not carried on under pressure of external necessity or danger. . . . It is in the spontaneous play of his faculties that man finds himself and his happiness. . . . Work and play here take on a different meaning, and become equivalent to servitude and freedom. The change consists in the subjective point of view from which the distinction is now made. We no longer mean by work all that is done usefully, but only what is done unwillingly and by the spur of necessity. By play we are designating, no longer what is done fruitlessly, but whatever is done spontaneously and for its own sake, whether it have or not an ulterior utility.[1]

Work, therefore, is what is imposed on us. Play is our 'free activity,' our 'holiday life.' Art then becomes either (*a*) an adornment, an idle pursuit, a frivolous matter at best (or in the deathless words of the New York City Board of Education of some three decades back, a 'useless frill'); or (*b*) the activity of the ideally free person, not hung up in the toils of daily necessity and survival; in either case, not essentially part of our lives as they are in fact lived. It is easy to meld the metaphors 'work' and 'technology' (for assuredly, both are highly metaphorical terms in the context of this discussion). 'Technology', according to the myth connotes 'technique' and *techne*, or skill. It separates the artisan from

the artist, the worker from the creator, the mass from the elite. Technology is what trade schools or vocational schools teach. And we all know who they are for. Indeed, the liberally educated man, and the child who is to become that man, need to 'know about' such things. So we go slumming, educationally, and include such peripheral and 'special' subjects as 'shop,' 'home economics,' and 'personal typing' – (not 'business typing,' for that already connotes a vocational track) – in our schools. The liberally educated person has to be democratic and 'well-rounded' – but not a mechanic!

Carry this mythology one step further: for there it separates 'fine art' from 'craft' or 'skill' even further, in a peculiar perversion of 'art': drafting or mechanical drawing is a technical subject. Drawing from the figure, or painting still lifes still has an air of 'rule' about it: there are judgments of better and worse, correct and incorrect to be made; and the last thing we want our spontaneous and free young 'artist' to be subjected to is the backward and reactionary notion that what he has drawn or painted is 'incorrect.' As long as it is an 'authentic' and 'imaginative' expression (of what? his 'feelings'? his 'plastic intuitions'? his 'colour sense'?), it makes its own rule. The child is genius in germ, giving the rule to art. The only 'art,' which suits this image of the spontaneous creator is a spontaneous 'art,' a use of the imagination unhampered by the strictures of craft, of rule, of tradition, of 'better' and 'worse.' ("That's *fine*, Cecily – isn't that *beautiful*! Did you make it up all *by yourself*?")

To be sure, this fairy tale about permissiveness in child art has its grim complement in the rigidities and aesthetic idiocies of the 'color-inside-the-line' school, or the 'color-the-grass-green-and-the-sky-blue' school, or the 'trace carefully' school. It is socially revealing that the permissive anti-art practices abound in upper-middle-class schools, and that the repressive anti-art practices abound in working-class and ghetto schools. But they are complementary practices, for both reinforce, and are reinforced by the myth, with its false separation of 'necessity' and 'spontaneity.' To be sure, there are exceptions, and there is a body of intelligent art-education theory, and a noble corps of intelligent and sensitive teachers of art who struggle daily with the physical impossibilities of schools which lack facilities and supplies, and with the curricular and intellectual impossibilities which demand either that art be taught in an entirely 'vocational' way; or that it be consigned to the limbo of its disutility; or that it satisfy doting parents who just love all

that creativity as a release from the burdensome constraints of the 'real' subjects. (It ranks with gym and 'free-play' periods, since, as we all know, art is a release of tension.)

Granting the dominance of the more repressive modes of art-activity, one needs to separate them from the legitimate demand, in art, for discipline and for skilled craftsmanship. For the myth-engendered version of freedom and spontaneity in art is equally perverse in at least two respects.

One perversion is that what is being demanded of the child is something he or she cannot deliver; and yet, the assumption is that it is *always* being delivered, on every occasion. Spontaneity, one of the greatest of human *achievements*, has been cheapened into randomness; and thereby, freedom into license. The child has been cheated, because achievement is so cheap.

A second perversion socially more degenerate is that the child's attitude toward (and aptitude for) disciplined work, as an avenue of delight and satisfaction in achievement, has been sold short, cut short, impugned. The requirements of craft, of skill, the acquisition of the historical and cultural achievements of the species, have been denigrated as tired requirements imposed by a dead traditionalism. However, the social effect of 'liberation' from this tradition is *not* the production of a humanistic or aesthetic elite, but rather of a generation of humanistic and aesthetic illiterates; young people robbed of their potentiality for creativity, in whatever field, by the myth that creation is easy: all you have to do is do it!

The social context of this view of art and of art education is not the art world itself. Here, struggle with ideas, with the limits of skill and style, with the medium *and* with the message, is daily and fierce. No joy without pain there. Rather, the social context of what I would call the 'arty' view of art (its dilettantism aside) is the separation of hand and head, of feeling and intellect, of work and leisure, of production and creation. And thus, because work and the workaday unfortunately *do* have the connotations Santayana ascribes to them, in our society – they constitute, most often, activity under constraint, done because it must be, for the sake of survival – the infelicity of work, its dullness, its lack of spiritual life, its sheer and weighty necessity are all burdens. And in this image of work, the workaday – and therefore also, the 'technological' – is framed.

Art then becomes escapist. It becomes Sunday art, holiday art,

museum art, divorced from life and from the social and personal demands of life. And by 'life,' here, I mean no grim prison of daily needs, but *whatever* it is that is the ongoing, self-preserving activity of a man, or a society. To the extent that the mass of men are required to become mere production machines, in order to 'preserve their being,' to that extent a society has become dehumanized, as the cost of its self-preservation; and for this, there are social and political remedies, one hopes. Yet to surrender human labor, work, production of the very means of one's existence to the level of dull routine, or to work as a necessary evil, seems to admit that life itself can be preserved only at the expense of the spirit; or worse yet, that one person's spiritual freedom from toil can be purchased only at the expense of another person's labor; that human exploitation is the condition for the freedom of art.

This social and ideological consequence, it seems to me, is at one with the separation of art and technology, as 'free' and 'slavish' activities. Technology is assigned to the precincts of toil – it is for 'them,' for those who have no recourse; whereas art is for 'us,' by virtue of our enlightenment – and more, by virtue of our leisure. And if we ourselves are caught in the daily grind ('technology') then our saving grace, our escape, our 'other world' is art. Art becomes the dream-world, the heaven of the imprisoned spirit; and therefore, the separation of art from real life becomes absolute: not merely acknowledged, but accepted and reinforced. The model of art as freedom, in a world and a life caught in the constraints of necessity, is therefore, at best, an escapist model; far from being a vision of a world-to-be-achieved, it is assigned the status of a vision of what real life can never be like, and of what we may enjoy only on the margins and in the rare moments of felicity in our ordinary lives. It is pie in the sky. The myth thus puts art beyond reach – an interlude, in the 'free periods' of our learning, sandwiched between serious subjects like mathematics and social studies, and good for letting off steam, or as therapy for disturbed children, or for providing the 'regular' teachers with a coffee break (at the expense of 'messing up the room' – for despite its exalted status as the realm of free creation, art is seldom accorded the literal room which 'necessary' subjects deserve). The separation of 'art' and 'technology' therefore acts to preserve and enhance quietism and passivity: an acceptance of the 'unaesthetic' conditions of a workaday life, as inevitable, since that's what 'technology' inevitably is like.

Worse yet: having fostered an illusion concerning art, we lose the

reality of technology as well. For one thing, we lose (and lose for our children) the breathtaking realization of the capacity of people to produce their own existence, to support and maintain life by their labor, their inventiveness, their capacity for social cooperation, their plain hard work. We give up the conceptual and ideological importance of the dignity of labor, in the face of its perversion by an often repressive industrial-technological social organization. We forget that human beings making things for human use, even in the context where the only aim of production seems to be either sheer survival, or sheer profit from the work of others. And so, we forget that people become human, by means of their work, just as they may become dehumanized by it. Human skill, artisanship, craft, technology, – the human ability to shape means to ends, by deliberate intelligence and imagination – is not something we can afford to surrender. Yet on the mythic view of technology as repressive, regimented activity, it is surrendered. In a world desperately in need of human productivity and skill, we can hear complaints about the 'work-ethic,' and mewlings about 'greening' and 'self-liberation.' Where it is plain that we cannot survive without technology, it becomes a mark of almost insane self-privilege to insist that we dump it. Such attitudes are, in effect, the marks of privilege of an elitism which holds itself beyond the social needs of a struggling mankind. At best they are the marks of petit bourgeois despair over the possibilities of effective social change. The other side of this despair is cynicism and escapism – the same escapism which leads to a perversion of the concept of art.

   This is not to deny the importance of a critique of an alienated technology – one which has lost its basis in the humanism of labor as the productive activity by which human needs are satisfied. There are social roots for this abstraction of technology from its human contexts, certainly. Technology has become an instrument for repression. It has been used for conquest, for inhumane ends, for the profit and power of the few. It has produced mega-death, to order. Political and military abuses of technology aside, the division of labor has gone so far in complex industrial societies like our own that we each, simply, produce practically nothing that we need, but contribute only to that collective social production which generates anonymous products for use. Indeed, one of the appeals of art is that its products are often autographic, uniquely identified with personality and individuality, in the face of technological anonymity. Yet, the solution which the myth proposes is to

kill the goose because the golden egg has been stolen from us. Instead of an effort to regain technology for human ends, it is condemned *in toto*, and abstractly. Those who have benefited most from it – in terms of relative affluence, or freedom from dire need – are often most anxious to persuade others to surrender it. But this seems to me more a cry of impotence, in the face of failure to control and use technology for human welfare, than a rational and militant program for regaining control over it.

What eventuates is a disdain for technology, and with it, a disdain for labor and craft (except in the hothouse or drawing-room versions of a quaint or primitive artsy-crafty activity, or in the escapist versions of commune greeneries, where 'doing your own thing' is reminiscent of nothing so much as Marie Antoinette's 'milkmaids' in the pastoral dairy, set up in the midst of a suffering, prerevolutionary France). This disdain for technology, for labor, for craft, expresses itself most precisely in the misconstrual of technology, and in the 'slumming' attitude toward vocational education.

The translation of this cultural myth into alternative and conflicting models of education is not so blatant as to be obviously self-condemning. In fact, dressed up in the language of 'freedom' versus 'repression,' this confrontation of art and technology cannot help but win our liberal sympathies. The rhetoric of the myth, in educational contexts, is persuasive enough, if we swallow the primary error in the characterizations of art and technology themselves. Very simply, and schematically, the cultural myth would translate educationally thus: deemphasize all rote learning, all the inherited disciplinary forms, all the quasi-military features of the school. Deemphasize or eliminate the merely 'vocational.' Let the child flower in the exercise of a free imagination, unconstrained by backward notions of grading, of the 'correct' and the 'incorrect.' Let spontaneity rule, for the child must be free to discover what he or she can or will. Let the child be an 'artist' in all of his educational enterprises, playing with the materials and resources, rather than being trained in their use. At best, simply have these resources available so that the child can choose them in the course of his free and creative exploration. Eliminate the punch-in-punch-out tyranny of the clock, the seating plan, the lesson-plan. Let the teacher be another such available resource, not directing, or guiding, but rather enhancing the child's own motivations. Eliminate the conceptions of disciplined work, of a schedule of achievements, of a linear scale of

required 'progress'; eliminate the training of 'skills' – let them rather be acquired 'naturally' where needed (as in learning to walk). School is not 'work' – it is play. We are not producing trained automatons, but creative agents. Not every child has to learn the same thing – rather, let the child learn at will what he or she wills. You can lead a horse to water etc., etc.

What, by contrast, does the cultural myth see as a technological model of education? Again, very schematically: find out what the needs of society are. Fit the educational program to the most efficient satisfaction of these needs. Screen children for their vocations, either by merit or by social class (no matter, the quotas will be fulfilled). Thereafter, train, condition, direct, mold so that the proper products will be produced for the needs at hand. The curriculum requires literacy specific to the tasks to be performed by the 'products' in their eventual vocations. Teach the discipline needed for 'real life' later on; obedience, perseverance, responsiveness to reward and punishment. Teacher says what's what. Authority reigns.

The technological model is, in many essentials, Plato's, in the *Republic*. I present it in its horrow-show version here just to be fair to the spirit in which it is rejected.

To say that these are the models of education which the myth proposes is not to say that such models exist only in the mythic imagination. They exist in fact. The myth has its embodiments. Certainly the permissive and the repressive exist side by side, in real life; they are perfectly compatible. They are two sides of the same coin, and no conflict or struggle ensues as long as each remembers which side of the coin it is on. And it is this separation, in real life, this abstraction and one-sided existence of the permissive and the repressive which is reflected in the myth. That the separation is 'false' is a normative claim for the unity and complementarity of the aesthetic and the practical, of freedom and necessity. But the 'false' separation does exist, and can be abided by a society which needs its art to be escapist and its technology to be exploitative. The myth doesn't invent the division; it dignifies it, and in this case, under the guise of a liberal humanism. What is at issue here is not simply the rosy model of a 'free' educational environment, nor the grim model of the school-factory, as instrument of society and authority. Rather, it is the misidentification of the first as somehow based on the model of 'art,' and of the second, on the model of 'technology.' What is at issue is whether some normatively adequate conception of art and

technology could entail such educational interpretations. The answer is 'no'; and therefore, the conceptions of art and technology which do entail such interpretations are normatively inadequate. That is to say, they are wrong. Were one to argue that 'normative adequacy' begs the question, and that the question is not what art and technology ought to be, but what they in fact are in this time and place, I would argue that 'art' and 'technology' are normative terms to begin with, and that it depends upon whose norms one is willing to adopt.

Since the issue lies in the dissection of the cultural myth, let me proceed to an alternative account of art and technology, not as conflicting models of education, but as complementary ones.

First, what positive features are there to art as a model for education? Art, among other things, is an exploration; an exploration into the properties and potentialities of a medium, first; for color, shape, movement, sound, language itself, are malleable, manipulable, and resistant, and the exploration of their possibilities is a piece of juvenile education without which fingers, eyes, ears, and limbs will remain underdeveloped. Moreover, the medium offers scope to the imagination. For art is, second, an exploration of the possibilities of the imagination. It is a search for precise expression, for adequacy in the presentation and representation of a feeling, an idea, an image, Moreover, since images do not come pre-formed in the mind, only *then* to be objectified, art concerns the process of forming or creating ideas or images, as plastic or visual or aural constructions. And so it is, at the same time, an exploration of modes and capacities one has both for creating and communicating such feelings, ideas, images.

The model for education is clear enough, at two levels at least: *First, that the practice of art, the exploration which art affords, is a necessary part of the education of any human being*. Without it, the capacities of people for technical skill and empathetic understanding of the materials of their world will remain cramped and debased. For the materials of their world are all the materials of their human praxis. These include words, sounds, and colors, as well as wood, steel, plastic; the medium of expressive gesture, as well as the medium of the visual image, or the vivid phrase. So, at one level at least, art is the model for education in the simple sense that a proper education includes aesthetic exploration, not as a frill or adornment, not as an escape, or a therapy, but as a necessity for the education of the child's sensibility to the world. In this sense, education should simply include what is necessary for human life – for a

good life, in which abilities and sensibilities are most fully realized; and so, art proper, beyond the school, is simply the highly perfected expression of such a human need. What art includes in its perfected expression, is therefore a model for what education should include, as an introduction into a fully human life – or at least, into the expectation for such a life, and the appreciation of it which will help make it worth struggling for.

Second, art is a model for education in a more analogical, less direct way, *insofar as it involves delight, or enjoyment, or pleasure in its achievements*. But this is not simply an argument for the values of pleasure; rather, it involves the social education of the child to the community of shared feeling, shared delight. The appreciation of art is a profound mode of social feeling, of the appreciation of other people. This is a subtle matter. For it is not simply that in viewing a great work of art, or hearing great music, we are awed by a Michelangelo or a Beethoven. We are rather awed by the art work itself; we delight in it, exercise and expand our sensibilities in the experience of it, enlarge our intellectual scope by the demands it makes on our understanding. All this is response to the art work itself, and not to its label, or its reputation. Yet, in the aesthetic experience of the work of art, we are in concourse with others; more than that, we are shaped, in our sensibilities, and in the modes of our response, to the cultural inheritance of the species, across time and space and in a communion as profoundly human as the sexual or the religious. In a word, we are humanized by great art; and can become dehumanized by base art. The lure of our enjoyment takes us beyond ourselves, just as love and faith do, and helps to make us species-beings, not mere atoms in a crowd. Thus, at two levels, as an activity which involves us with the materials of human *praxis*, and as a mode of humanizing and socializing our responsiveness to the human community, art provides a very general model for education.

Now in what ways does technology provide an alternative or even conflicting model for education? I confess I do not know. For what I refuse to accept is the dichotomy, in the first place. Suppose we construct a minimal model of the 'technological,' as having to do with means, applications, the 'how' of things. We take technology to be related to science, ordinarily, and not to art (except as 'industrial art' or 'craft'). In relation to science, technology presumably has to do with the embodiment, or the carrying out in practice, of principles or truths established 'theoretically' in the sciences. Landing on the moon is

therefore, simply, the application of Newtonian physics – the distances and speeds being too small to require any Einsteinian considerations of a relativistic sort; and too big to involve quantum-theoretical consider-ations. The technology, at least insofar as trajectories, masses, and forces are concerned, is simply a matter of engineering an embodiment of a theoretical idea already established. The landing is regarded as a feat of applied theory, an engineering achievement, not a theoretical one. (We discount the auxiliary technology and theory of fuels, of television transmission, of optical and radio guidance systems, etc. in this account). But surely, this view of the relation of technology and science is simpleminded! Granted that there is established theory, and that the engineering applications are not devised either as tests of the theory, nor as explorations of alternative theories; it is still clear that new theory, in a number of scientific fields, grows out of just such contexts of application. Further, apart from serendipity, scientific discovery is so closely linked to the development of instrumentation, to the technological possibilities opened up by new hardware and engineering techniques, and to the host of new problems which crowd in upon the solution of old ones, that the sharp division of scientific theory proper from technology may be retained only as a formal distinction, and not a distinction in practice. One may grant with impunity the valuable distinction between 'mission-oriented' scientific research and 'basic' research, and even insist that a good deal of scientific theorizing goes on in a way detached from contexts of solution to given practical problems (though not detached from the context of solutions to given theoretical problems). The relations between theory and practice in science *and* in technology are complex and subtle. For all this, in the living practice of science, the neat formal divisions between the theoretical and the practical become fuzzy, the membrane becomes highly permeable, and especially so in periods of active scientific-technological development.

  To press the point: if one were to separate the technological so discretely from the theoretical, in science, and then to insist on its further separation from the free and imaginative intellect on the one hand, and from the creative aesthetic sensibility on the other, the caricature left of technology would be thin indeed. Yet, one may argue, it is the case, right or wrong; that in actual practice, in actual social and production contexts, this division is enforced. Some are creators, some mere imitators, or mechanics. The division of labor *has* reduced many human activities to *mere* technology, to uncreative, rote repetition of a

procedure, devoid of any but the meanest skill, and certainly devoid of any intellect or sensibility. And, it will be argued, this is what we mean by 'technology' – the brute function of effecting what someone else has prescribed, no matter what skill is required.

But there is a not-so-subtle shift here, from the idea of carrying theory into practice, to the idea of sheer, mechanical, untheoretical practice, which violates the very sense of technology proper. That technology can become alienated is no more astounding than that theoretical science and art itself can become alienated. There is no special virtue of art or of science that can prevent them from becoming dehumanized; and no special vice of technology, that leads to its dehumanization. It is quite clear (both historically, and at present) that fine art and theoretical science can be dehumanized very effectively; and equally clear that technology can become humanized. What this means is that the sense of the human – the need for wholeness, for socially just and humane ends, for joy in work – is not prescribed or proscribed by the limits either of art, or science, or technology. What we require is that theory and practice, discovery of new truths and the deployment of old ones, imagination and skill, be closely linked to each other, so that their interaction ensures their mutual integrity.

What has this to say of a technological model for education? Simply that if we seperate the requirements of craft, skill, and application from the requirements of free, imaginative exploration, we have doomed the imagination to impotence, and skill to brute and blind activity. The effect of this is to leave the 'creative' either as an effete and ineffectual fantasy life; or worse, to create an elitist and hierarchial ideology, in which the 'creative' person is forever provided with the assurance that others, less fortunate, will carry out his ideas in practice, will do the 'dirty work.' In educational contexts, this connotes a disdain for practice, for sustained work, for the sense of a collaboration among equals. Does this mean that there is *no* division of labor, no special talent, no leadership, expertise, genius? I think not. Rather, it means that an alienating division of labor should not be imposed on our children in their education; for the 'creative' child can become just as alienated in his alleged 'creativity,' suffering from lack of means to make his free imagination count for something, as can the 'backward' child, assigned to his 'vocational' tasks.

The disdain for technology has another effect, related to these: the workaday, the technical, the applied is itself a domain of creativity and

freedom. The child knows this clearly, before he becomes corrupted by an elitist ideology. He can transform a task into a game, a skill into an exploration; his fantasy and imagination see possibilities in the ordinary, and he is magical with it. Work and wonder are not yet at odds in the child. The child's insight is profound here, though not self-conscious. We are responsible to recapture it in a self-conscious way, to nourish it and recognize it, and to discover (or rediscover) for ourselves this unity of play and work, of free vitality and disciplined form. Technology, on this model, is not the alternative, or conflicting model to that of art, but is part of what a viable model of art contains; conversely, the aesthetic, the explorative, the creative is what a viable model of technology contains. Far from conflicting, these are complementary models of a unified mode of human activity.

What, then, is politically and educationally vicious about the counterposition of these two models? Not simply their counter-position, nor even the distinctions one may draw between their relative emphases. Rather it is the uses and interpretations to which such counterposition gives rise. It is the false, exclusive and mythical models of art and technology which are vicious. They are vicious because they accept and harden the separation of two requisite and complementary modes of human activity; in doing so, the two modes are abstracted and seen as exclusive of each other. We invest, in these abstracted models, the very abstraction and alienation of those activities which plague our social and economic life. That is to say, we embody in these models a falseness which is *objective* in our own social life – the human falseness of a division between creation and work, between free human agency and slavish toil. Instead of liberating work, labor, technique, from its sheer instrumentalism, we denigrate it as slavish, and pretend that we can transcend its slavishness by eliminating it altogether. In favor of what? Why, art of course. But such an art would be devoid of its social role, its seriousness, its demands upon skill, craft, and discipline. It would be a pretense at art, not the living practice of it.

Far from being a liberating conception, this contrast of the two modes accepts the status quo and accommodates it. In doing so, this contrast encourages social attitudes in our children which tend to denigrate not only the technological, or the technically productive activity, abstractly, but the people who are engaged in it. In short, what is encouraged is a disdain for working people; and in the case of working-class students, a self-disdain, and disdain for their parents. Similarly, the contrast of the

two modes encourages admiration for those freed from the necessities of toil, for the creative spirits who have risen above the dull necessities of life. Unfortunately, these few are dream figures; they have escaped the system, or have dominated and used it. Such mythic models (of the 'artist,' presumably) are our contemporary versions of the 'star,' of the folk hero who beats the system and 'makes it' on his own. And our contemporary version of this myth has the same social function as did the Hollywood version of a generation ago: the goal is not to change or improve the system, but to beat it. The solution remains individual and not social. And the social and political effect is to displace real social change with a fantasy of self-achievement.

People need art; and people need technology. The world needs to be made aesthetic, humanized for mankind. But the conquest of war, of disease, of hunger demands the education of a generation technologically capable of effecting such changes, and morally and aesthetically committed to them. The cultural myth which disarms us in the face of such needs must be put to rest.

NOTE

[1] George Santayana, *The Sense of Beauty: Being the Outline of Aesthetic Theory,* Dover, New York, 1955, pp. 17–19.

ART AS HUMANIZING PRAXIS*

[1976]

I

Human beings become human in coming to know themselves as human. I will argue, in this paper, that the creation and appreciation of art – the very activity or praxis of art – is a praxis which comes to know itself, i.e. which takes *itself* as its own object; and that this very activity is a fundamental mode of human self-knowledge. It is therefore a humanizing praxis. The thesis I am rejecting is that art is a representation of human nature, a mirror for man, insofar as it represents semblances or images of human beings, or of actions, or of events or scenes *in general*. One standard view is that art is an imitation of nature, that it is a portrayal or depiction of something beyond art. This is certainly true, of representational art (whatever the canons of representation may be) but it is not in such portrayal or depiction that human self-knowledge develops, and thus, it is not in this sense that art is a humanizing praxis. What I will claim is that art – or the artwork – is a representation of a mode of action which is distinctively human – namely, the creation of artworks; in short, that art represents its own process of coming into being and insofar, exemplifies and objectifies the distinctively human capacity of creation. It is in the self-recognition of this creative capacity that human beings come to know themselves as human, in the specific sense that they come to know themselves as creators or as artists. Thus it is not *what* is portrayed, or depicted which provides the humanizing content of the artwork, but rather the reading back of the very process of its genesis which makes the artwork an objective representation of human creativity. Art thus exemplifies or symbolizes the activity of art. The artist thus becomes a model of the potentialities of human nature, of human creativity, in which the appreciator realizes his own species-capacity for creation, not necessarily in himself, but in another. Feuerbach already suggests such a view of the artwork. In his discussion of the great cathedrals as edifices ostensibly dedicated to God, he sees a self-celebration by the builders of their own craftsmanship: what is objectified in the cathedral is the capacity, the vision, the imagination

and the skill of the master-builders; their glorification of their own human abilities, embodied in a public form. One may elaborate the analysis further: insofar as the cathedral represents a vision and the ability to embody it, it represents the social capacity to support this vision in the public support for such construction, and in the manifold social labor embodied in the final product. Such a view of art as a mode in which human praxis recognizes itself in its own objects, and comes to realize the capacities and limits of human creativity in this objectified form, is a particular case of that reflexivity by which the dialectical tradition characterized the distinctively human. The injunction "know thyself" becomes more than an injunction. It becomes a definition of the human. The human is that form of being which can come to know itself, and to know itself in those characteristic modes which then define humanity. There is therefore no external or *a priori* 'human nature' but only that which is constituted in this activity itself, and the praxis or action whereby such self-knowledge is acquired becomes, therefore, humanizing praxis.

There is one other sense in which art is a mode of self-knowledge, at a secondary level: namely that in which a self-knowledge already achieved exemplifies itself or is represented *to* itself in the artwork. In this secondary sense, the artwork provides the reflective (or refractive) condition of *critical* self-knowledge. The phenomenon, well known to painters, in which, when one works without a model, the portrait takes on characteristic features of the painter's physiognomy, is, I think a case in point of this *dioptric*.

In order to develop this view further, let me set out some programmatic points which relate this account to the theme of this conference: the crises in Western culture.

(1) First, just as the activity of art is a humanizing activity, the repression and distortion of this activity, its elimination from the ordinary life world, or its retention in its false and alienated form leads to dehumanization, to an aesthetic deprivation whose result is a pathological social and psychological life.

(2) The commodification of art – the corruption of the activity of art into a process of production of commodities for exchange, and into an instrumentality for the facilitation of exchange – leads to just such a dehumanization.

(3) The role of art – of artistic creation and appreciation – has been corrupted in this way, both in terms of the market-place and the

forum – i.e. as commodity or as ideology – and there has developed a corruption of taste, for the mass of the public, whose symptoms are either the denigration of art, or the capacity to be satisfied with the lowest products of art, with *ersatz*, *kitsch* and vulgarity; and to require ever novel stimulation as shock-treatment for a degraded and depressed sensibility for art.

(4) This crisis in art is not an internal crisis of the artworld alone but is symptomatic of a deeper social and historical crisis in which sensibility itself has lost its objective function as a mode of self-knowledge by aesthetic means, and has become instead an abstract, and crude instrumentality of response for the sake of some other end – profit, ideological persuasion, or sheer distraction: *panem et circenses* as a safe catharsis for social and personal frustration and misery.

Before I develop this characterization of art as a humanizing praxis further, I wish to make two digressions: first, to examine the dialectical model of self-knowledge in general, of which the case of art is a particular one; second, to draw an analogy between the creation and appreciation of artworks, and Marx's account of the production process, and of the objectification which praxis yields as a mode of self-knowledge.

Let me therefore repeat the opening sentence in this context:

II

Human beings become human in coming to know themselves as human. This peculiar reflexivity of human nature was first deeply recognized by Hegel, who saw it as the very condition of consciousness itself – namely, that consciousness takes itself as its own object, and becomes fully self-conscious – i.e. fully human – consciousness in the recognition of itself in its own otherness, or in its objectification in another. The humanist interpretation – or transformation – of Hegel's phenomenology was most fully realized in Feuerbach's work – especially in the masterpiece of his middle period, *The Essence of Christianity*. Here, Feuerbach ascribed the specific character of being human to the capacity to know oneself as a species-being; i.e. to recognize an other as a *Thou*, as a being like oneself. This conscious recognition of the other as the condition of the self-recognition of one's species-nature, one's *gattungswesen* Feuerbach saw as the secret of religious consciousness: Human beings

come to know themselves in the alienated form of an other which is both like oneself – i.e. individual, suffering, mortal, dependent – and yet wholly unlike oneself – i.e. universal, impassible, immortal and wholly self-dependent, or *causa sui* – in short, *divine*. Therefore, Feuerbach argued, human beings come to know themselves as species-beings in the mystified form of an *image* of their own nature, which they take to be an independent reality beyond themselves – i.e. in the image of God. *Homo Homine Deus est*: Man is a God to Man. Feuerbach's demystification, or demythisation consisted, then, in revealing the image as image, and human being as its source or original.

In the dialectical tradition, then, the very condition of humanization is the representation of one's own nature as an object of consciousness, albeit in the form of an other. The fullest form of human existence is the self-conscious, and clear recognition of oneself *in* the representation, i.e. of the 'other' as oneself; but no longer as a merely finite individual ego, but as a social being, whose very existence and nature is dependent on others like oneself.

This abstract phenomenological humanism ascribes the species character of human beings to their self-recognition as human beings. It is abstract in the sense of Marx's critique of Feuerbach: the 'image' is of human being as such, and not in social and historical context; it is phenomenological in that it centers on the constitution of the image by a subject, and the subject's self-recognition in the relation of consciousness of the image. Feuerbach took religion to be the paradigm of this psychology and phenomenology of self-representation, though in its mystified form. His materialism consisted in the recognition of sensibility, need, feeling as the fundamental existential features of human life, and therefore as the basis for the representation of human nature. Sensibility here is the direct activity of the senses themselves, in revealing the material world as the reality in which human life is lived, and of which it is itself a product. Human needs are therefore not simply the needs of consciousness as such for an *other*, but rather, the needs which sensate, bodily, material existence requires for its fulfillment. Thus sensibility is already a representation of the genesis of human existence in nature, and of the continuing dependency of this existence upon nature. Feuerbach's later works – what one may characterize as his materialist humanism – translate the phenomenology of religious consciousness into the self-conscious naturalism which sees consciousness as the product of a dialectic of sensibility, i.e. as

itself the reflex of material human existence – i.e. sensate, needy, dependent existence. It is in the dialectic of this sensibility with its other that sensibility and feeling – *aistheisis* – develops as cognition or intellection – as *episteme*. The 'other', in which human beings come to recognize themselves as human, is no longer simply the *Thou* of religious consciousness, but the natural world itself insofar as it becomes a world-for-us – a world either designed or made to meet the needs of human existence. It is this transformation of the world into a resource for human existence that makes of it the mirror of our needs, and thereby, the representation, in this form, of our species-nature or our essence. It is thus in this humanization of nature that human beings come to be human – the humanized world becomes the *other* whereby humanity achieves itself. Other human beings enter into this representation, then, not merely as intuitively recognized *Thous*, but insofar as they meet the needs of our own existence. It is our need for others, as conditions of our own existence, that makes them the objects of our feeling, our love, our dependency; and it is this, in turn, which characterizes our nature as essentially social, or as a species-nature.

It was Marx, of course, who recognized most clearly these implications of Feuerbach's work, and who therefore was also able to see how far short Feuerbach's own theory fell of the requirements for a concrete account of human needs and of the activity or *praxis* which was required to meet them. What Marx saw as Feuerbach's failing was the delimitation of the need-satisfying activity to the realm of reflective or contemplative activity of consciousness, or at most, of feeling and sensibilitity. In short, it is *this* characterization of human nature which falls short of the richness and complexity of concrete human existence – i.e. its historical articulation, in terms of the development of historical modes of *praxis*. Young Marx (like young Hegel before him) recognized that the fundamental activity of meeting needs was a *practical* activity of transforming the world, and not merely its transformation in the understanding, or in conscious reflection, or in 'abstract' sensibility or feeling. Marx's eleventh thesis on Feuerbach, concerns 'changing the world' by the concrete intervention of what Marx called 'sensuous-concrete praxis' or 'revolutionary praxis'. In effect, the *praxis* to which Feuerbach's account is limited is the *praxis* of belief, of feeling, of consciousness. Thus, whereas Marx characterizes all praxis as a *conscious praxis*, Feuerbach's remains a praxis *of consciousness*, albeit not of purely reflective or intellective consciousness, but of sensibility, feeling, belief.

We are all familiar with the sequel, and with Marx's argument for a descent from the critique of the Holy Family to a critique of the earthly family – to the critique of society, of the state, of law, of political economy.

Ane yet the phenomenological elaboration of this critique – the development of a theory of the conscious reflection of world-transforming praxis – remained fragmentary in Marx and Engels, and has only recently begun to develop in its full philosophical form.

Lukács, in the early, pre-Marxist Heidelberg aesthetics (of 1911–1914) began to elaborate some of the phenomenological aspects of a distinctive form of this 'reflective praxis' in the domain of art. Marxist aestheticians, like Max Raphael, Ernst Fischer, Lucien Goldmann, elaborated similar ideas, as did some aestheticians of the Frankfurt school of neo-Marxism, notably, Walter Benjamin, Theodore Adorno and Herbert Marcuse.

The fundamental idea which runs through all this work is an old one; its power lies in the richness and concreteness of the elaboration which it makes possible. It is this: human beings come to know themselves (as human) in form of the objects or artifacts which they fashion, and in which human nature comes to represent itself to itself. The simple form of the thesis is: 'Art is a mirror for man'. But the Marxist emphasis is not simply on the contemplation of oneself in the mirror of art: rather, it is in the transformative activity of art itself – of the praxis of art – that human self-knowledge develops. Thus, it is not the product – the artwork, the completed and dead image – which is the mirror of human nature, but rather the process of artistic creation itself, *and* the process of *re*creation in the act of aesthetic appreciation which constitutes the active humanizing praxis of art.

We have arrived, by a detour, at the beginning of this paper then: its subject is art as humanizing praxis. But now, its context should be clear as well.

There is, however, a second detour I wish to make; and this one will appear even more digressive than the first. The first detour was for the sake of establishing the context of a dialectical phenomenology in which the work of art and the process of artistic creation and appreciation would be understood as a process of human self-objectification for the sake of human self-recognition. The second detour is concerned with establishing what I take to be a striking parallel between Marx's account of the productive process, in his critique of political economy, and what I take to be a Marxist account of the creative process in art. It is this:

Marx's critique of political economy starts with a fundamental critique of classical political economy, in particular, of its theory of value. In place of the classical theory – i.e. that the value of a commodity is determined by its cost of production – Marx introduces his distinctive form of the labor theory of value, his theory of surplus-value. It should be familiar enough so that I will not repeat it here in any elaborate form, but only in the briefest sketch:

(1) The only source of value, the only creator of value, is 'living' labor – i.e. the active labor of human beings in transforming natural resources into objects of use to meet human needs. This activity is conscious, teleological praxis.

(2) In capitalist production, this 'living' labor itself becomes a commodity, and is sold by the worker at a price: what he sells is not the *product* of his labor, but rather his labor-power (at a certain rate for a certain time). The price of this commodity is the *wage*, which, according to Marx, is determined in the long run (through market fluctuations) by the cost of reproducing that labor power.

(3) Surplus value is that portion of the value produced by living labor which exceeds the cost of reproducing the labor power which was required to produce it. It is, as Marx says, unpaid labor, and it is possible precisely because living labor *creates* value, and creates it as new value, beyond the value required to produce or reproduce the labor power itself.

(4) 'Dead' or 'congealed' labor is the surplus-value accrued by the capitalist – and transformed by him into instruments of production: machinery, plants, etc. The transformation of the surplus value produced by 'living' labor into its 'dead' or 'congealed' form, i.e., into capital, is what makes capitalist production possible, and what makes the growth or accumulation of capital possible.

(5) This accumulated capital however has no value in itself, but only insofar as this 'congealed' value is released or transferred by living labor – by the work-process – into commodities which can be exchanged. Surplus value, then, does not come from the use of capital, or of machinery, etc., but only from creative living labor: and *past* surplus value, accrued in the form of 'dead' labor – e.g. machinery, etc. – is released into the marketplace for exchange only by the hand of living labor.

This brief sketch has certain features which provide a striking analogy to the creative process in art. The artwork, as 'product', i.e., as completed object, has no value in itself, except insofar as it is the

congealed value of the creative process itself; and it is 'released', so to speak, only by the hand of creative appreciation – i.e. a reproduction of the process of its creation. Now it is anomalous to speak of the process of creative work in art as a 'production' process leading to a commodity exchange; just as it is anomalous to speak of the process of creative appreciation of the artwork as if it were either *merely* a recapitulation or imitation of the process of creation, or as if it were a process of 'consumption' of art. It is true that some aestheticians – e.g. John Dewey – speak of creation and appreciation of art as analogous to production and consumption of commodities. But what is the aesthetic analogy to *exchange*? That the analogy quickly breaks down is obvious. So too, it is obvious that to constrain the process and product of artistic creation to the model of commodity production in capitalism is to reduce art to a mode of capitalist production, which is absurd. (Except in the real-life situation in the artworld, where this absurd phenomenon has become a fact of life, and art is produced for exchange as a commodity, and exists only in this alienated form.)

Still, the model has a heuristic value, for in it, the ongoing creative activity or praxis by which human beings produce and reproduce their existence is understood as an objectification – (a *vergegenständigung*) or an externalization (an *entaüsserung*) – in which the activity of creation of values to satisfy needs comes to be represented or embodied in the product of that activity. It is in this sense that the commodity-form of the product of capitalist production *represents* the *activity* of capitalist production, and embodies in itself the production-relations of capitalism – i.e. the relations of exchange; and it is in the interpretation of the apparent or phenomenal form of this relation that Marx discovers or reveals what is hidden by this appearance – namely, that the relation is one of exploitation, and that this exploitation of living labor is defined by the production of surplus value. Thus Marx reads back from the *object* – from the objectified or externalized form of this production and exchange process – the reality which it represents. The reality – the social relations of production under capitalism – is thus revealed in its objective form, in the commodity – and is demystified. For, according to Marx, the apparent or phenomenal form of the commodity represents the relation between capitalist and worker, between exploiter and exploited as if it were *merely* an exchange relation, an external contractual relation among equals, in the form of civil society, or of the marketplace. But this is the 'mystified' or alienated form of the relation: its image taken to be

the reality. The mistake of taking the image for the social reality which underlies it is the same mistake which Feuerbach alleges against religious consciousness: of taking the religious image as an independent reality, instead of as the reflection or representation of the human reality of which it is a projection or an objectification. Thus, Marx talks of this 'mistake' as commodity fetishism, in completely Feuerbachian language, in the famous passage in Capital.

A commodity is therefore a mysterious thing, simply because in it the social character of men's labour appears to them as an objective character stamped upon the product of that labor; because the relation of the producers to the sum total of their own labor is presented to them as a social relation, existing not between themselves, but between the products of their labour . . . a definite social relation between men, assumes, in their eyes, the fantastic form of a relation between things. In order, therefore, to find an analogy, we must have recourse to the mist-enveloped regions of the religious world. In that world the productions of the human brain appear as independent beings endowed with life, and entering into relation both with one another and the human race. So it is in the world of commodities with the products of men's hands. This I call the Fetishism which attaches itself to the products of labour, so soon as they are produced as commodities, and which is therefore inseparable from the production of commodities. This Fetishism of commodities has its origin, as the foregoing analysis has already shown, in the peculiar social character of the labour that produces them.[1]

The analogy to art comes in Marx's (and Engels') characterization of the labor process as the very process by which human beings achieve their humanity and their sociality. In the *Paris Manuscripts*, in *the German Ideology* and elsewhere in the early works, Marx and Engels already characterized this process of humanization, though in negative critical terms: if, in fact, it is by labor – by the production and reproduction of the conditions for continued human existence – that man 'makes himself'; if human being is, in effect, the product of creative human praxis, then man's essence as self-creating being becomes lost to him, if the very control over this process and its product is taken from him – i.e. if the very activity which characterizes and generates his humanity becomes alienated, or estranged (*entfremdet*). It is only in the reappropriation of this control over their essential humanizing activity, that human beings overcome this de-humanizing alienation, and reachieve their humanity. The particular form of capitalist production is, according to Marx, an *essentially* alienated form of this essential human activity or praxis, and therefore is dehumanizing. Revolutionary praxis, then, is the means to the overcoming of alienation, and the reappropriation, or generation of fully human life.

But the overcoming of this alienation comes only with the reappropriation of the production process – of the creative or value-producing activity of self-sustenance – as one's own – and not as one's own individually, but under the conditions of social ownership of the means of production, i.e. under socialism.

III

The previous discussion remains abstract with respect to the activity of art itself – it sets forth an ontological characterization of the artwork as the embodied, externalized or objectified form in which creative or value-producing human praxis becomes reflective, comes to know itself in the form of an other. It remains to make the analysis more concrete with respect to the artforms themselves.

The premise here is that the artwork has no value in itself – neither as a use-value, i.e. an aesthetic value; nor as an exchange value, i.e. as an alienated object of exchange, as a commodity – but has its value only in its 'living' form – in the process of creation or of appreciation. This is not a matter of 'attitude' towards the *art object*, as some theories of aesthetic attitude claim – Rather the attitude is only the symptom of an underlying function – of the aesthetic activity itself. One does not place oneself 'in' an attitude; rather one engages in a mode of praxis, of which the 'attitude' is a function. This is true not only of the artist and of the appreciator, but of the critic as well. The *praxis* of criticism elaborates another mode in which human capacities (i.e. critical capacities) come to be known by means of the *praxis* of criticism itself. It is these modes of creation, appreciation and criticism (or aesthetic judgment) which come to humanize us precisely in that we *become* artists, appreciators or critics – and thus become human.

To be human, then, is to be capable of creation, appreciation, criticism. There is no *a priori* essence of humanity, then, which these forms of human action exemplify: rather, this essence is self-constituted by the praxis itself, and exactly to the degree that this praxis becomes self-conscious. The condition of self-consciousness here is the existence of such objects which not only embody the activity or praxis of art, but in which this praxis is reappropriated or recognized as one's own.

The artforms – whether in literature or poetry, in the visual arts, in music, in dance, in the aesthetic praxis of everyday life – become such objectifications. When the activity becomes ritual or automatic; when

the object comes to be seen only in its surface appearances – e.g. as depiction or portrayal, as thematic content, or even as sheer aesthetic surface (to use Prall's phrase), or as form alone – the human content of the artwork becomes transparent and redundant: it is seen through, but not realized. In this case, one may speak of an alienated aesthetic consciousness, a fetishism of the artwork, in which the object is taken as an autonomous and independent reality. Even here, if my thesis is correct, the objectification of a mode of human action is realized, but in a mystified form – as if it were a non-human, or trans-human reality. In such a case, there is an un-self-conscious relation to one's own capacity for art, and a denigration of the existential need for art. If the human being is that which comes to know itself as human, then in this case of alienated aesthetic consciousness, such self-knowledge is not achieved, such humanization does not take place.

Yet, there is another form of this alienated consciousness. It is exemplified in this very discourse itself, as a *philosophical* discourse. Suppose we agree that it is only in the self-conscious reappropriation of one's own artistic capacities that we are humanized by art. This *philosophical* conviction is however not itself a praxis of art, but rather of reflection. It is *about* art, but it is not artistic praxis. It reveals to us another aspect of our humanity – our reflective, philosophical capacities. So too, criticism is not itself artistic praxis (except insofar as either a philosophical or critical discourse is a work of art – most are not, and this one is certainly not). Thus, the peculiar reflective self-consciousness of theoretical or critical praxis humanizes us as well, but is not to be confused with the praxis of art. The substitution of art-theory for aesthetic praxis, the *mistake* of taking a critical or reflective philosophical praxis for the praxis of art *also* fetishizes – it does not fetishize the artwork, but takes something – the reflective critical consciousness *about* art and its praxis – in place of art. It is a form of idolatry, therefore.

What then is the non-idolatrous, non-fetishistic, non-alienated praxis of art: it is ultimately the activity of creation and recreation itself – it is the actual participation in art – the imperative to which becomes a humanistic imperative in concrete social terms.

Non Art, anti-art, etc., is an irony based on a recognition of alienation: it takes the artwork, and the artist as themselves the *subject* matter of art, and in this sense it is a perverse recognition of the alienation of the traditional forms of art. It says: Look at me, at what *I'm* doing – not at the product of art *except* insofar as it reveals my creative activity.

Traditional art has become opaque: its human content and genesis has been lost in the layers of traditionalized, ritualized and alienated response to artworks. The newer artforms focus on a return to the process: but perversely: Duchamps, by making his signature definitive of the artwork; Dada, by violently making us aware of the *choice* which is made in artistic practice: by choosing the strange, exotic and unusual in order to reveal the *act* of choice as dominant in art. Even the dialectically perverse *disappearance* of the artist in anonymous art is a *protest* against this loss, rather than an affirmation of it. The insistence on an impersonal and anonymous objectivity – a focus on the object – simply asserts in a painful way, which makes us attend to it, – the alienated condition of art: as such it is a negative critique – a cry of despair. And as we know, despair is an explicit cathexis – a way of accommodating the conditions of repression without breakdown. Institutionalized despair, in the artworld, becomes a passive adaptation to the status quo: a mode of accepting the unacceptable. It is a way of preserving one's humanity at least in the mode of non-self-rejection, of non-denial; but not yet in the spirit of *affirmation*: an elaboration of sensibility.

The humanizing praxis of art is not, of course, simply that it makes us happy, or enhances our sensibilities, or realizes our capacities for self-conscious praxis. Art acquaints us as well with tragedy, with despair, with the recognition of our suffering as human suffering, as anguish and grief structured not simply as cries and groans, but as expressions of shared and communicable feelings. It is critical and reflective praxis even in these negative modes because our fears and pains become objectified to us in their human configurations. We recognize the look, the affect, and we give it form. In all this, there is more than the feeling involved – there is the feeling recognized and articulated either in its ritual modes – e.g. in the rituals of grieving for the dead, which constitute an artform of social praxis – or in its specifically 'artistic' forms, in literature, painting, music, etc.

Art therefore humanizes us as much in our capacity for tragedy as for joy; and the forms of our creative activity even here are historical and social forms, for our sensibility is not animal but human, the product of our life with and through others like ourselves, and within those structures of everyday practice which we create and which create us. We thus come to know ourselves in all this richness and historical complexity through art, insofar as it represents, in their elaboration and development, all these modes of our own praxis.

NOTES

* This paper was delivered at the meeting of the International Husserl and Pheno-menological Research Society, March 28, 1974, Montreal, Canada.
[1] K. Marx, *Capital: A Critique of Political Economy*, Moscow: Foreign Languages Publishing House, 1959, p. 72.

NAME INDEX

371

SUBJECT INDEX

375

Inert 93ff, 97f, 107
Organic
  Diderot's Account of 322f
  LaMettries' Account of 319f
  Maupertuis' Account of 318f, 320f, 327f
  Sensibility in 116, 321f
  Self–Active 92ff
Maturation, Theory of 158
Maturational Sequence 171f
Meaning xx
  Analysis 19f
Mechanism, Mechanistic
  See also Materialism
  and Life 321ff, 331f
  Newtonian 298
  Reduction 297
Memory, Spinoza's Model 241f
Mental, Demarcation of 165
Mental Entities 13, 165
Mental Events 104, 117n5
Mental Operations 165
Mental Pictures xxi
  See also Internal Representations
Mentalism 107, 164ff
Mentalistic Description as Unscientific 108
Metaphor, Metaphorical xix, 10, 218
  See also Imagination
Metaphysical
  See also Ontological
  Commitments 24, 28ff, 36
  Constructs 163
  Ideas 91
  Questions 90
  Terms 44, 46ff, 73
  Theories 55f, 59
  Thesis, in Diderot 337n2
Metaphysics 24ff, 27, 40ff, 90ff, 277
  See also Dualism, Materialism, Mind-Body, Monism
  Adequacy of 77f
  and Choice of Research Problems 61
  Classical 21, 53f, 163
  and Common Sense 81ff
  Continuous with Science 71ff, 83
  as Critical Device 71
  Definition of 65f
  Demarcation of 57ff, 63f, 90

'Doing –' 86
Elimination of 90f
  as Esoteric 36f
  and Explanation 40, 50
  as Formal System 76f
  Greek 99
  as Heuristic for Science xxiv, 36, 39, 40ff, 46ff, 61f, 67ff, 84f, 91f, 102, 151, 163, 231ff, 253
  History of in Science 48ff, 67
  Idealist 90, 297
  Irrationalist 91
  Justification of 80ff
  and Language 83f
  and Logic 77
  and Models 28, 51, 69
  as Model of Reality 72, 77, 85
  and Myth 59
  and Rationality 70f
  and Science 35, 40ff, 68f, 90ff, 130, 231
  Speculative 90
  and Stories 73ff
  and Understanding 70
  and the Universal 82f
Method
  Cartesian 49
  Scientific 54
  Spinoza's 243ff, 262f
Methodology
  and Ontology 37
  and 'True Search' 158
Mimesis 24
Mimicry 201
Mind xxi, 236, 240, 259
  See also Materialism, Matter, Mind-Body
  as Active 107
  as Experimental Apparatus 102
  Operations of 165
  as Social-Historical 117
  Structures of 157f, 166, 167f, 169, 193
Mind-Body
  See also Materialism, Matter, Mind
  Dualism 99, 104, 193, 234
  as Interactive 250
  Materialist Theory of 251f
  Monism 250
  Problem 104ff
    and Common Sense 111f

SYNTHESE LIBRARY

Studies in Epistemology, Logic, Methodology,
and Philosophy of Science

Managing Editor:
JAAKKO HINTIKKA, (Academy of Finland, Stanford University
and Florida State University)

Editors:
ROBERT S. COHEN (Boston University)
DONALD DAVIDSON (University of Chicago)
GABRIËL NUCHELMANS (University of Leyden)
WESLEY C. SALMON (University of Arizona)

15. C. D. Broad, *Induction, Probability, and Causation. Selected Papers.* 1968, XI + 296 pp.
16. Günther Patzig, *Aristotle's Theory of the Syllogism. A Logical-Philosophical Study of Book A of the Prior Analytics.* 1968, XVII + 215 pp.
17. Nicholas Rescher, *Topics in Philosophical Logic.* 1968, XIV + 347 pp.
18. Robert S. Cohen and Marx W. Wartofsky (eds.), *Proceedings of the Boston Colloquium for the Philosophy of Science 1966-1968,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume IV. 1969, VIII + 537 pp.
19. Robert S. Cohen and Marx W. Wartofsky (eds.), *Proceedings of the Boston Colloquium for the Philosophy of Science 1966-1968,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume V. 1969, VIII + 482 pp.
20. J.W. Davis, D. J. Hockney, and W. K. Wilson (eds.), *Philosophical Logic.* 1969, VIII + 277 pp.
21. D. Davidson and J. Hintikka (eds.), *Words and Objections: Essays on the Work of W. V. Quine.* 1969, VIII + 366 pp.
22. Patrick Suppes, *Studies in the Methodology and Foundations of Science. Selected Papers from 1911 to 1969.* 1969, XII + 473 pp.
23. Jaakko Hintikka, *Models for Modalities. Selected Essays.* 1969, IX + 220 pp.
24. Nicholas Rescher *et al.* (eds.), *Essays in Honor of Carl G. Hempel. A Tribute on the Occasion of His Sixty-Fifth Birthday.* 1969, VII + 272 pp.
25. P. V. Tavanec (ed.), *Problems of the Logic of Scientific Knowledge.* 1969, XII + 429 pp.
26. Marshall Swain (ed.), *Induction, Acceptance, and Rational Belief.* 1970, VII + 232 pp.
27. Robert S. Cohen and Raymond J. Seeger (eds.), *Ernst Mach: Physicist and Philosopher,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume VI. 1970, VIII + 295 pp.
28. Jaakko Hintikka and Patrick Suppes, *Information and Inference.* 1970, X + 336 pp.
29. Karel Lambert, *Philosophical Problems in Logic. Some Recent Developments.* 1970, VII + 176 pp.
30. Rolf A. Eberle, *Nominalistic Systems.* 1970, IX + 217 pp.
31. Paul Weingartner and Gerhard Zecha (eds.), *Induction, Physics, and Ethics: Proceedings and Discussions of the 1968 Salzburg Colloquium in the Philosophy of Science.* 1970, X + 382 pp.
32. Evert W. Beth, *Aspects of Modern Logic.* 1970, XI + 176 pp.
33. Risto Hilpinen (ed.), *Deontic Logic: Introductory and Systematic Readings.* 1971, VII + 182 pp.
34. Jean-Louis Krivine, *Introduction to Axiomatic Set Theory.* 1971, VII + 98 pp.
35. Joseph D. Sneed, *The Logical Structure of Mathematical Physics.* 1971, XV + 311 pp.
36. Carl R. Kordig, *The Justification of Scientific Change.* 1971, XIV + 119 pp.
37. Milič Čapek, *Bergson and Modern Physics,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume VII. 1971, XV + 414 pp.

38. Norwood Russell Hanson, *What I Do Not Believe, and Other Essays* (ed. by Stephen Toulmin and Harry Woolf), 1971, XII + 390 pp.
39. Roger C. Buck and Robert S. Cohen (eds.), *PSA 1970. In Memory of Rudolf Carnap*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume VIII. 1971, LXVI + 615 pp. Also available as paperback.
40. Donald Davidson and Gilbert Harman (eds.), *Semantics of Natural Language.* 1972, X + 769 pp. Also available as paperback.
41. Yehoshua Bar-Hillel (ed.), *Pragmatics of Natural Languages.* 1971, VII + 231 pp.
42. Sören Stenlund, *Combinators, λ-Terms and Proof Theory.* 1972, 184 pp.
43. Martin Strauss, *Modern Physics and Its Philosophy. Selected Papers in the Logic, History, and Philosophy of Science.* 1972, X + 297 pp.
44. Mario Bunge, *Method, Model and Matter.* 1973, VII + 196 pp.
45. Mario Bunge, *Philosophy of Physics.* 1973, IX + 248 pp.
46. A. A. Zinov'ev, *Foundations of the Logical Theory of Scientific Knowledge (Complex Logic)*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume IX. Revised and enlarged English edition with an appendix, by G. A. Smirnov, E. A. Sidorenka, A. M. Fedina, and L. A. Bobrova. 1973, XXII + 301 pp. Also available as paperback.
47. Ladislav Tondl, *Scientific Procedures*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume X. 1973, XII + 268 pp. Also available as paperback.
48. Norwood Russell Hanson, *Constellations and Conjectures* (ed. by Willard C. Humphreys, Jr.). 1973, X + 282 pp.
49. K. J. J. Hintikka, J. M. E. Moravcsik, and P. Suppes (eds.), *Approaches to Natural Language. Proceedings of the 1970 Stanford Workshop on Grammar and Semantics.* 1973, VIII + 526 pp. Also available as paperback.
50. Mario Bunge (ed.), *Exact Philosophy – Problems, Tools, and Goals.* 1973, X + 214 pp.
51. Radu J. Bogdan and Ilkka Niiniluoto (eds.), *Logic, Language, and Probability. A Selection of Papers Contributed to Sections IV, VI, and XI of the Fourth International Congress for Logic, Methodology, and Philosophy of Science, Bucharest, September 1971.* 1973, X + 323 pp.
52. Glenn Pearce and Patrick Maynard (eds.), *Conceptual Change.* 1973, XII + 282 pp.
53. Ilkka Niiniluoto and Raimo Tuomela, *Theoretical Concepts and Hypothetico-Inductive Inference.* 1973, VII + 264 pp.
54. Roland Fraïssé, *Course of Mathematical Logic – Volume 1: Relation and Logical Formula.* 1973, XVI + 186 pp. Also available as paperback.
55. Adolf Grünbaum, *Philosophical Problems of Space and Time.* Second, enlarged edition, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XII. 1973, XXIII + 884 pp. Also available as paperback.
56. Patrick Suppes (ed.), *Space, Time, and Geometry.* 1973, XI + 424 pp.
57. Hans Kelsen, *Essays in Legal and Moral Philosophy*, selected and introduced by Ota Weinberger. 1973, XXVIII + 300 pp.
58. R. J. Seeger and Robert S. Cohen (eds.), *Philosophical Foundations of Science. Proceedings of an AAAS Program, 1969*, Boston Studies in the Philosophy of

Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XI. 1974, X + 545 pp. Also available as paperback.

59. Robert S. Cohen and Marx W. Wartofsky (eds.), *Logical and Epistemological Studies in Contemporary Physics*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XIII. 1973, VIII + 462 pp. Also available as paperback.

60. Robert S. Cohen and Marx W. Wartofsky (eds.), *Methodological and Historical Essays in the Natural and Social Sciences. Proceedings of the Boston Colloquium for the Philosophy of Science 1969-1972*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XIV. 1974, VIII + 405 pp. Also available as paperback.

61. Robert S. Cohen, J. J. Stachel and Marx W. Wartofsky (eds.), *For Dirk Struik. Scientific, Historical and Political Essays in Honor of Dirk J. Struik*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XV. 1974, XXVII + 652 pp. Also available as paperback.

62. Kazimierz Ajdukiewicz, *Pragmatic Logic*, transl. from the Polish by Olgierd Wojtasiewicz. 1974, XV + 460 pp.

63. Sören Stenlund (ed.), *Logical Theory and Semantic Analysis. Essays Dedicated to Stig Kanger on His Fiftieth Birthday*. 1974, V + 217 pp.

64. Kenneth F. Schaffner and Robert S. Cohen (eds.), *Proceedings of the 1972 Biennial Meeting, Philosophy of Science Association*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XX. 1974, IX + 444 pp. Also available as paperback.

65. Henry E. Kyburg, Jr., *The Logical Foundations of Statistical Inference*. 1974, IX + 421 pp.

66. Marjorie Grene, *The Understanding of Nature: Essays in the Philosophy of Biology*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XXIII. 1974, XII + 360 pp. Also available as paperback.

67. Jan M. Broekman, *Structuralism: Moscow, Prague, Paris*. 1974, IX + 117 pp.

68. Norman Geschwind, *Selected Papers on Language and the Brain*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XVI. 1974, XII + 549 pp. Also available as paperback.

69. Roland Fraïssé, *Course of Mathematical Logic – Volume 2: Model Theory*. 1974, XIX + 192 pp.

70. Andrzej Grzegorczyk, *An Outline of Mathematical Logic. Fundamental Results and Notions Explained with All Details*. 1974, X + 596 pp.

71. Franz von Kutschera, *Philosophy of Language*. 1975, VII + 305 pp.

72. Juha Manninen and Raimo Tuomela (eds.), *Essays on Explanation and Understanding. Studies in the Foundations of Humanities and Social Sciences*. 1976, VII + 440 pp.

73. Jaakko Hintikka (ed.), *Rudolf Carnap, Logical Empiricist. Materials and Perspectives*. 1975, LXVIII + 400 pp.

74. Milič Čapek (ed.), *The Concepts of Space and Time. Their Structure and Their Development*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XXII. 1976, LVI + 570 pp. Also available as paperback.

75. Jaakko Hintikka and Unto Remes, *The Method of Analysis. Its Geometrical Origin and Its General Significance*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XXV. 1974, XVIII + 144 pp. Also available as paperback.

76. John Emery Murdoch and Edith Dudley Sylla, *The Cultural Context of Medieval Learning. Proceedings of the First International Colloquium on Philosophy, Science, and Theology in the Middle Ages – September 1973*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XXVI. 1975, X + 566 pp. Also available as paperback.

77. Stefan Amsterdamski, *Between Experience and Metaphysics. Philosophical Problems of the Evolution of Science*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XXXV. 1975, XVIII + 193 pp. Also available as paperback.

78. Patrick Suppes (ed.), *Logic and Probability in Quantum Mechanics.* 1976, XV + 541 pp.

79. Hermann von Helmholtz: *Epistemological Writings. The Paul Hertz/Moritz Schlick Centenary Edition of 1921 with Notes and Commentary by the Editors.* (Newly translated by Malcolm F. Lowe. Edited with an Introduction and Bibliography, by Robert S. Cohen and Yehuda Elkana), Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XXXVII. 1977, XXXVIII+204 pp. Also available as paperback.

80. Joseph Agassi, *Science in Flux*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XXVIII. 1975, XXVI + 553 pp. Also available as paperback.

81. Sandra G. Harding (ed.), *Can Theories Be Refuted? Essays on the Duhem-Quine Thesis.* 1976, XXI + 318 pp. Also available as paperback.

82. Stefan Nowak, *Methodology of Sociological Research: General Problems.* 1977, XVIII + 504 pp.

83. Jean Piaget, Jean-Blaise Grize, Alina Szeminska, and Vinh Bang, *Epistemology and Psychology of Functions*, Studies in Genetic Epistemology, Volume XXIII. 1977, XIV+205 pp.

84. Marjorie Grene and Everett Mendelsohn (eds.), *Topics in the Philosophy of Biology*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XXVII. 1976, XIII + 454 pp. Also available as paperback.

85. E. Fischbein, *The Intuitive Sources of Probabilistic Thinking in Children.* 1975, XIII + 204 pp.

86. Ernest W. Adams, *The Logic of Conditionals. An Application of Probability to Deductive Logic.* 1975, XIII + 156 pp.

87. Marian Przełęcki and Ryszard Wójcicki (eds.), *Twenty-Five Years of Logical Methodology in Poland.* 1977, VIII + 803 pp.

88. J. Topolski, *The Methodology of History.* 1976, X + 673 pp.

89. A. Kasher (ed.), *Language in Focus: Foundations, Methods and Systems. Essays Dedicated to Yehoshua Bar-Hillel*, Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XLIII. 1976, XXVIII + 679 pp. Also available as paperback.

90. Jaakko Hintikka, *The Intentions of Intentionality and Other New Models for Modalities.* 1975, XVIII + 262 pp. Also available as paperback.

91. Wolfgang Stegmüller, *Collected Papers on Epistemology, Philosophy of Science and History of Philosophy*, 2 Volumes, 1977, XXVII + 525 pp.
92. Dov M. Gabbay, *Investigations in Modal and Tense Logics with Applications to Problems in Philosophy and Linguistics.* 1976, XI + 306 pp.
93. Radu J. Bogdan, *Local Induction.* 1976, XIV + 340 pp.
94. Stefan Nowak, *Understanding and Prediction: Essays in the Methodology of Social and Behavioral Theories.* 1976, XIX + 482 pp.
95. Peter Mittelstaedt, *Philosophical Problems of Modern Physics,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XVIII. 1976, X + 211 pp. Also available as paperback.
96. Gerald Holton and William Blanpied (eds.), *Science and Its Public: The Changing Relationship,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XXXIII. 1976, XXV + 289 pp. Also available as paperback.
97. Myles Brand and Douglas Walton (eds.), *Action Theory. Proceedings of the Winnipeg Conference on Human Action, Held at Winnipeg, Manitoba, Canada, 9-11 May 1975.* 1976, VI + 345 pp.
98. Risto Hilpinen, *Knowledge and Rational Belief.* 1979 (forthcoming).
99. R. S. Cohen, P. K. Feyerabend, and M. W. Wartofsky (eds.), *Essays in Memory of Imre Lakatos,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XXXIX. 1976, XI + 762 pp. Also available as paperback.
100. R. S. Cohen and J. J. Stachel (eds.), *Selected Papers of Léon Rosenfeld,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XXI. 1978, XXX + 927 pp.
101. R. S. Cohen, C. A. Hooker, A. C. Michalos, and J. W. van Evra (eds.), *PSA 1974: Proceedings of the 1974 Biennial Meeting of the Philosophy of Science Association,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XXXII. 1976, XIII + 734 pp. Also available as paperback.
102. Yehuda Fried and Joseph Agassi, *Paranoia: A Study in Diagnosis,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume L. 1976, XV + 212 pp. Also available as paperback.
103. Marian Przełęcki, Klemens Szaniawski, and Ryszard Wójcicki (eds.), *Formal Methods in the Methodology of Empirical Sciences.* 1976, 455 pp.
104. John M. Vickers, *Belief and Probability.* 1976, VIII + 202 pp.
105. Kurt H. Wolff, *Surrender and Catch: Experience and Inquiry Today,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume LI. 1976, XII + 410 pp. Also available as paperback.
106. Karel Kosík, *Dialectics of the Concrete,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume LII. 1976, VIII + 158 pp. Also available as paperback.
107. Nelson Goodman, *The Structure of Appearance,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume LIII. 1977, L + 285 pp.
108. Jerzy Giedymin (ed.), *Kazimierz Ajdukiewicz: The Scientific World-Perspective and Other Essays, 1931 - 1963.* 1978, LIII + 378 pp.

109. Robert L. Causey, *Unity of Science.* 1977, VIII+185 pp.
110. Richard E. Grandy, *Advanced Logic for Applications.* 1977, XIV + 168 pp.
111. Robert P. McArthur, *Tense Logic.* 1976, VII + 84 pp.
112. Lars Lindahl, *Position and Change: A Study in Law and Logic.* 1977, IX + 299 pp.
113. Raimo Tuomela, *Dispositions.* 1978, X + 450 pp.
114. Herbert A. Simon, *Models of Discovery and Other Topics in the Methods of Science,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume LIV. 1977, XX + 456 pp. Also available as paperback.
115. Roger D. Rosenkrantz, *Inference, Method and Decision.* 1977, XVI + 262 pp. Also available as paperback.
116. Raimo Tuomela, *Human Action and Its Explanation. A Study on the Philosophical Foundations of Psychology.* 1977, XII + 426 pp.
117. Morris Lazerowitz, *The Language of Philosophy. Freud and Wittgenstein,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume LV. 1977, XVI + 209 pp.
118. Tran Duc Thao, *Origins of Language and Consciousness,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx. W. Wartofsky), Volume LVI. 1979 (forthcoming).
119. Jerzy Pelč, *Semiotics in Poland, 1894 - 1969.* 1977, XXVI + 504 pp.
120. Ingmar Pörn, *Action Theory and Social Science. Some Formal Models.* 1977, X + 129 pp.
121. Joseph Margolis, *Persons and Minds, The Prospects of Nonreductive Materialism,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume LVII. 1977, XIV + 282 pp. Also available as paperback.
122. Jaakko Hintikka, Ilkka Niiniluoto, and Esa Saarinen (eds.), *Essays on Mathematical and Philosophical Logic. Proceedings of the Fourth Scandinavian Logic Symposium and of the First Soviet-Finnish Logic Conference, Jyväskylä, Finland, 1976.* 1978, VIII + 458 pp. + index.
123. Theo A. F. Kuipers, *Studies in Inductive Probability and Rational Expectation.* 1978, XII + 145 pp.
124. Esa Saarinen, Risto Hilpinen, Ilkka Niiniluoto, and Merrill Provence Hintikka (eds.), *Essays in Honour of Jaakko Hintikka on the Occasion of His Fiftieth Birthday.* 1978, IX + 378 pp. + index.
125. Gerard Radnitzky and Gunnar Andersson (eds.), *Progress and Rationality in Science,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume LVIII. 1978, X + 400 pp. + index. Also available as paperback.
126. Peter Mittelstaedt, *Quantum Logic.* 1978, IX + 149 pp.
127. Kenneth A. Bowen, *Model Theory for Modal Logic. Kripke Models for Modal Predicate Calculi.* 1978, X + 128 pp.
128. Howard Alexander Bursen, *Dismantling the Memory Machine. A Philosophical Investigation of Machine Theories of Memory.* 1978, XIII + 157 pp.
129. Marx W. Wartofsky, *Models: Representation and Scientific Understanding,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XLVIII. 1979 (forthcoming). Also available as a paperback.
130. Don Ihde, *Technics and Praxis. A Philosophy of Technology,* Boston Studies in

the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XXIV, 1979, XIV + 152 pp. Also available as a paperback.

131. Jerzy J. Wiatr (ed.), *Polish Essays in the Methodology of the Social Sciences,* Boston Studies in the Philosophy of Science (ed. by Robert S. Cohen and Marx W. Wartofsky), Volume XXIX. 1979, XIV + 260 pp. Also available as a paperback.

132. Wesley C. Salmon (ed.), *Hans Reichenbach: Logical Empiricist.* 1979, X + 782 pp.

133. R.-P. Horstmann and L. Krüger (eds.), *Transcendental Arguments and Science.* 1979, VIII + 314 pp. Also available as a paperback.